



Chromatin accessibility profiling by ATAC-seq

Fiorella C. Grandi^{1,2,3}, Hailey Modi^{1,2,3}, Lucas Kampman^{1,2,3} and M. Ryan Corces^{1,2,3}✉

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) provides a simple and scalable way to detect the unique chromatin landscape associated with a cell type and how it may be altered by perturbation or disease. ATAC-seq requires a relatively small number of input cells and does not require a priori knowledge of the epigenetic marks or transcription factors governing the dynamics of the system. Here we describe an updated and optimized protocol for ATAC-seq, called Omni-ATAC, that is applicable across a broad range of cell and tissue types. The ATAC-seq workflow has five main steps: sample preparation, transposition, library preparation, sequencing and data analysis. This protocol details the steps to generate and sequence ATAC-seq libraries, with recommendations for sample preparation and downstream bioinformatic analysis. ATAC-seq libraries for roughly 12 samples can be generated in 10 h by someone familiar with basic molecular biology, and downstream sequencing analysis can be implemented using benchmarked pipelines by someone with basic bioinformatics skills and with access to a high-performance computing environment.

Introduction

Mapping alterations in cell states is a key aspect of understanding biological systems. Whether in development, differentiation or disease, cell state is governed by changes in gene expression that are, in turn, orchestrated by changes in gene regulatory programs. In recent years, it has become increasingly clear that these gene regulatory programs are established and controlled by the activity of transcription factors (TFs) that both interpret and alter the underlying epigenetic state of chromatin. The epigenetic state of chromatin can be regulated by a variety of mechanisms, including chemical modification of both DNA and histone proteins that, in turn, alter chromatin dynamics and high-dimensional chromatin structure. We now recognize that chromatin can exist in several different states^{1,2} that are defined by combinations of different epigenetic modifications and are associated with particular gene regulatory patterns. At the two ends of the spectrum are (i) active gene regulatory elements such as enhancers, promoters and insulators, which are bound by DNA-binding proteins, and (ii) inactive regions of silenced or poised chromatin, which are generally refractory to gene expression machinery³. Understanding the epigenetic state of chromatin in a certain biological context can shed light onto the molecular mechanisms underlying the observed gene expression patterns.

The epigenome can be assayed in a variety of ways, ranging in specificity. The most specific methods assay the location and abundance of a particular histone modification, DNA modification or TF using an antibody-based pull-down method and refinements of the classical chromatin immunoprecipitation with sequencing (ChIP-seq) technique⁴. Antibodies against the target of interest are used to enrich genomic segments from bulk chromatin. These marked fragments are then prepared into a library, sequenced using high-throughput sequencing and aligned to the genome. Genome regions where many reads overlap appear as peaks, indicating that the target of interest was present at that particular locus in a large proportion of the cells. Several refinements have been made to this classic technique^{5,6}. Recently, the sensitivity of these antibody-based techniques to determine DNA–protein interactions has been improved with the development of chromatin immunocleavage techniques (ChIC)⁷, including CUT&RUN⁸ and CUT&TAG⁹, which tether a nuclease or Tn5 transposase to the targeting antibody via the use of Protein-A. These modifications increase the resolution of the resulting protein binding site, remove the need for an immunoprecipitation step and decrease the amount of input material required. These techniques, and other derivations of similar technologies, including ChIPmentation¹⁰, CoBATCH¹¹, chromatin integration labeling sequencing (ChIL-seq)^{12,13} and single-cell (sc)ChIC-seq¹⁴, were recently adapted to work on single cells^{11,15–17}.

¹Gladstone Institute of Neurological Disease, San Francisco, CA, USA. ²Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA.

³Department of Neurology, University of California San Francisco, San Francisco, CA, USA. ✉e-mail: ryan.corces@gladstone.ucsf.edu

Such targeted profiling methods can provide key insights into the epigenetic changes driving a particular cell state but require a priori knowledge of the expected mechanism.

In some cases, it can be more helpful to obtain a broader picture of the gene regulatory landscape, particularly when a phenomenon is observed but the specific nature of the epigenetic changes remains unknown. For this reason, alternate strategies have been developed to probe the gene regulatory landscape by mapping all TF binding sites, agnostic to the precise TF identities, for a cell type or tissue. A variety of such agnostic profiling techniques have been developed, including deoxyribonuclease I hypersensitivity sequencing (DNase-seq)^{18–20}, formaldehyde-assisted isolation of regulatory elements (FAIRE-seq)²¹, micrococcal nuclease digestion with sequencing (MNase-seq)^{22,23} and nucleosome occupancy and methylome sequencing (NOME-seq)²⁴, and were recently reviewed in detail²⁵. Among the first of these assays to be developed was DNase-seq²⁰, which applied high-throughput sequencing to the classic technique of DNase digestion to measure different chromatin states²⁶. The activity of DNase is obstructed at sites bound by TFs or nucleosomes, thus protecting these DNA fragments and allowing them to be identified via sequencing. Simultaneously, the binding of TFs creates adjacent nucleosome-free regions of DNA that are hypersensitive to enzymatic digestion by DNase. Thus, mapping of these hypersensitive sites has been used extensively to identify putative gene regulatory elements in an unbiased manner. In particular, DNase-seq has become the gold-standard technique for TF footprinting^{27–29}. Similarly, MNase-seq is based on the use of MNase, an endo-exonuclease that cleaves regions of DNA that are not protected by nucleosomes or DNA-binding proteins²³. It can be thought of as an orthogonal approach to DNase-seq in that it measures nucleosome-occupied regions and is most often used to map nucleosome occupancy quantitatively genome wide^{30,31}.

The original DNase-seq and MNase-seq assays traditionally had complex, time-consuming library preparation protocols and required large numbers of cells as starting material. To address some of these limitations, while keeping the agnostic profiling of chromatin, the assay for transposase-accessible chromatin using sequencing (ATAC-seq) was developed³². ATAC-seq uses the activity of an engineered, hyperactive Tn5 transposase³³ preloaded with sequencing adapters to determine the sites of accessible chromatin. The development of ATAC-seq was based on two observations: (i) a transposase had previously been used to generate ‘tagmentation’ libraries, in which a Tn5 transposase was preloaded with sequencing adapters and used to simultaneously fragment and tag genomic DNA for high-throughput sequencing library preparation³⁴ and (ii) the observation that *in vivo* Tn5 could efficiently insert into nucleosome-free regions³⁵. ATAC-seq generates genome-wide regulatory maps that are highly similar to those derived from DNase-seq and MNase-seq³², while reducing library preparation complexity and hands-on time. ATAC-seq has been widely adopted owing to its low input material requirements (<50,000 cells) and short processing time, which facilitates data generation from large numbers of samples. Here we describe an updated version of the ATAC-seq protocol, based on our previous Omni-ATAC protocol³⁶, and provide helpful guidelines and benchmarks to ensure high-quality data generation from most, if not all, input cell types.

Applications of ATAC-seq

ATAC-seq provides a simple and scalable way to assay the regions of the genome that are bound by TFs, and to compare how these landscapes change between particular contexts or perturbations. This is accomplished using *in vitro* transposition of sequencing adapters into native chromatin (Fig. 1a). Each unique transposition event, termed an ‘insertion’, marks a location in the genome where a Tn5 transposase dimer is able to access DNA and perform a cut-and-paste reaction. The transposase simultaneously fragments the DNA and inserts sequence handles that are then used for amplification during library preparation. A sequenceable ATAC-seq DNA fragment is created by two separate transposase insertion events (Fig. 1b). The precise biochemical interactions that govern Tn5 transposition at these sites are not yet fully understood. It is generally accepted that the binding of a TF to DNA is associated with the creation of adjacent nucleosome-free regions and that these are associated with increased Tn5 transposition. We discuss several nuances to interpreting the resulting Tn5-chromatin accessibility in Supplementary Note 1.

Extensive profiling efforts have shown that regions of Tn5-accessible chromatin can be found at promoters, located proximal to the transcription start site (TSS), and at intergenic regions of the genome largely corresponding to enhancers, insulators or silencers^{32,37,38}. These patterns and locations of Tn5 accessibility, especially those at distal elements, are often cell type or cell state specific³⁷. Thus, ATAC-seq represents a valuable tool to understand how cells control gene expression, by

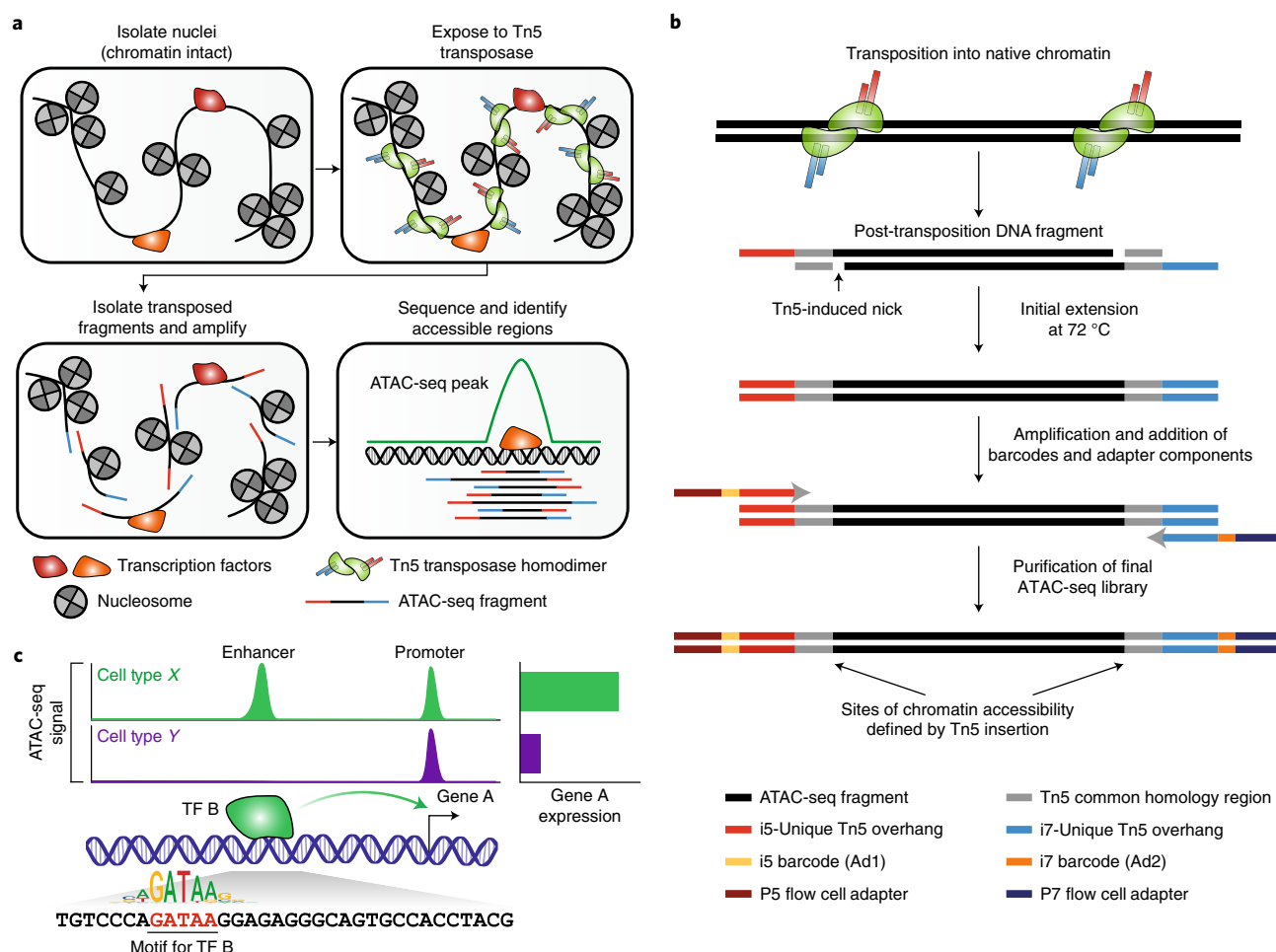


Fig. 1 | Schematic of the ATAC-seq transposition reaction and library preparation. a, Overview of the different steps in ATAC-seq. Nuclei are isolated from cells, maintaining the chromatin structure and any associated DNA-binding proteins, including nucleosomes and TFs, intact. This chromatin is then exposed to the Tn5 transposase, which acts as a homodimer to simultaneously fragment the chromatin and insert sequences containing PCR handles that enable downstream amplification with i5/P5 and i7/P7 ATAC-seq adapters. Only fragments that receive an i5/P5 adapter at one end and an i7/P7 adapter at the other end will be properly amplified and sequenced. After sequence analysis of the library fragments, genomic regions enriched for many Tn5 transposition events are designated as peaks of chromatin accessibility, or ATAC-seq peaks. **b**, Detailed schematic of the fragments generated by transposition into native chromatin in Step 10. After Tn5 insertion of the PCR handles, the nicks left behind by the transposase are filled in during the initial 72 °C extension in the first step of the barcoding PCR. Then, the fragments are barcoded and prepared for sequencing. **c**, A toy example of what hypothetical ATAC-seq data might look like. Cell types X and Y both have ATAC-seq peaks at the promoter of gene A. However, cell type X has a cell-type-specific enhancer upstream of the TSS, which drives higher levels of expression of gene A. In the ATAC-seq data, this would appear as a cell-type-specific ATAC-seq peak, which would be differentially accessible between cell types X and Y. In this toy example, the differential accessibility at the enhancer is driven by TF B, which binds to the cell type-specific enhancer and increases gene expression. ATAC-seq data can also help reveal the TF binding motif associated with such differential peaks, here depicted as 'GATAA'.

mapping the location of putative gene regulatory elements. After processing and alignment of ATAC-seq fragments, enrichment of Tn5 transposition events at specific genomic regions is used to identify peaks of Tn5-accessible chromatin in each sample. These are often termed 'ATAC-seq peaks'. Chromatin accessibility signal within these peak regions can be compared between different sample types using established pipelines (see the 'Data analysis' section of 'Experimental design'), and serve as the starting point for a variety of downstream analyses. For example, peaks can be linked to putative gene targets by using orthogonal chromatin conformation capture datasets or by naively assigning each peak to the nearest gene. These predicted gene regulatory interactions can provide a hint as to the functional importance of a given peak. Often, genes with several ATAC-seq peaks in their promoter and gene body are inferred to be actively expressed in that cell type. While gene expression is more accurately measured by RNA sequencing (RNA-seq), ATAC-seq can explain the mechanism behind how gene expression is regulated or why it might be different between two cell types or conditions.

A common application of ATAC-seq is to identify novel enhancers or gene regulatory regions for a given cell type or cell context of interest. For example, observing an ATAC-seq peak in a region 5 kb upstream of gene A's TSS in cell type X but not in cell type Y suggests that a cell-type-specific enhancer may regulate gene A's expression in cell type X (Fig. 1c). Such observations have been made comparing cells before and after exposure to a variety of stimuli^{39–41} or between different cell types or developmental timepoints^{42–46}. Differential activity of regulatory elements has been found in a wide array of diseases^{38,47–50}, and such findings motivate efforts to use ATAC-seq to characterize gene regulatory landscapes in large patient cohorts^{38,51,52}. ATAC-seq has also been used to fine-map disease-associated genetic variants identified through genome-wide association studies (GWASs), thus enabling prediction of putative functional noncoding alterations^{53–57}. Historically, GWASs have identified single-nucleotide polymorphisms (SNPs) that often reside in noncoding regions, making it difficult to interpret their functional impact. Intersecting regulatory regions identified through ATAC-seq in disease-relevant cell types with GWAS SNPs can help build hypotheses about which of those SNPs may affect gene expression and thus mediate disease-associated phenotypes.

ATAC-seq peaks can also be annotated for the presence of various TF motif sequences, and enrichment tests can be used to predict the drivers of differential chromatin accessibility. Such motif-based analyses can be useful for comparing different cell types or disease states, understanding differentiation or developmental trajectories^{45,58} or comparing how in-vitro-derived cells compare to their in vivo equivalents⁵⁹. In the toy example above, the cell-type-X-specific ATAC-seq peak 5 kb upstream of gene A might be bound by TF B, which is not expressed in cell type Y (Fig. 1c). Such a cell-type-specific difference in TF expression might result in many peaks throughout the genome being bound by TF B, leading to an enrichment of peaks harboring the motif for TF B within the subset of peaks that are differentially accessible between cell types X and Y. These types of analyses have been used to show changes in TF usage between different cell states—for example, during metastasis in small-cell lung cancer⁶⁰, before and after exposure to inflammatory stimuli⁴⁰ or during reprogramming of fibroblasts into neurons⁵⁹. ATAC-seq data can also be used to infer the positions of nucleosomes⁶¹, providing insights into chromatin regulation beyond TF binding. In summary, ATAC-seq is an effective technique for uncovering the gene regulatory changes that govern why cells express certain genes and how gene expression changes are mediated.

Comparison with other chromatin profiling methods

The diversity of available techniques to map DNA regulatory elements can make it challenging to determine which technique is the most appropriate and informative for a specific application. In Table 1, we compare some technical and experimental aspects of the most commonly used techniques for mapping DNA regulatory elements: ATAC-seq^{32,36}, DNase-seq¹⁹, MNase-seq²⁵, ChIP-seq⁵ and targeted CUT&TAG⁶², to help new users decide which assay is best for their particular application. We recommend basing this decision on (i) what kind of information is necessary to answer the specific research question and (ii) what kind of input material is available. In general, epigenomic profiling is appropriate to answer how or why a cell type or tissue might exhibit gene regulatory changes. For questions dealing predominantly with what is changing, we suggest beginning with RNA-seq⁶³.

To decide which epigenomic profiling technique is right for a specific application, the first decision to make is whether a broad or targeted approach is appropriate. If the question involves the location of a specific TF, DNA-binding factor or histone modification, TF ChIP-seq or related technologies would be most appropriate⁵. ChIP-seq of histone modifications such as H3K27ac, H3K4me1 or H3K4me3 has been widely used to map DNA regulatory elements across a variety of cell and tissue types. For example, H3K27ac predominantly marks active promoters and enhancers. Thus, by combining genome-wide maps of one or more of these modifications, an overall picture of the epigenome can be formed. However, in general, these assays are best interpreted in combination, and a single mark may not capture all types of DNA regulatory elements (i.e., promoters, enhancers and silencers). Alternatively, if the epigenomic profiling will predominantly serve for hypothesis generation or the desired information is better captured by measuring global patterns in regulatory element activity, techniques such as ATAC-seq, DNase-seq or MNase-seq may be more appropriate. These techniques have the advantage of being able to capture many different types of regulatory information with a single assay.

Compared with DNase-seq and MNase-seq, ATAC-seq has two major advantages for 'omics' users: (i) the library preparation is fast and easy with the ability to generate sequence-ready libraries

Table 1 | Comparison with other chromatin profiling protocols

	ATAC-seq	DNase-seq	MNase-seq	CUT&TAG or related ChIC techniques
Enzyme type	Tn5	Endonuclease	Endonuclease and exonuclease	Tn5 conjugated to an antibody via protein A
Sequence bias?	Yes; complex, Tn5 insertion bias, with preference for A/T in insertion site and C/G flanking ^{131–133}	Yes; complex, partially dependent on enzyme concentration and on methylation status of CpGs ^{85,134}	Yes; preferential cutting upstream of A/T compared with G/C ^{135,136}	Yes; dictated by antibody used to guide Tn5 and by Tn5 bias
Number of input cells or nuclei for standard bulk assays	500–50,000	1–10 million	10,000–100,000	100,000–500,000
Low-input/single-cell methods available?	Yes ^{86,87} ; commercial solutions available	Yes ⁶⁷	Yes ⁶⁶	Yes ^{62,64,137–139}
Sample type	Fresh or cryopreserved cells or nuclei. Fresh or frozen tissues	Fresh or cryopreserved cells or nuclei. Fresh or frozen tissues. Formaldehyde cross-linked or FFPE samples	Fresh or cryopreserved cells or nuclei. Fresh or frozen tissues. Formaldehyde cross-linked samples	Fresh or cryopreserved cells or nuclei. Fresh or frozen tissues
Library preparation time	~10 h for 12 samples (this protocol)	1–3 d	~2 d	1–2 d
Technical considerations	Library quality is highly dependent on cell viability. Protocol alterations are required for use on fixed cells, and data quality is often reduced for those samples	Enzyme concentration and digestion duration may need to be optimized to sample type	Enzyme concentration and digestion duration may need to be optimized to sample type	The amount of antibody used must be titrated for the cell type or sample. This will be a function of the strength of the antibody and the abundance of the target protein
Sequencing type	Paired-end	Size of fragments selected affects downstream analysis ^{2,8}	Apparent nucleosome occupancy is a function of MNase concentration	The assay is as specific as the primary antibody used. Additionally, this is a targeted technique, so additional libraries must be made of each modification or protein tested
Sequencing depth	Low: 10 million read pairs per sample with Omni-ATAC	Single-end Medium/high: 20–50 million uniquely mapping reads per sample; 200 million for TF footprinting	Single-end High; 150–200 million reads per sample (human) ¹⁴⁰	Single-end or paired-end Very low; 3 million read pairs per sample
Data produced	Tn5-accessible chromatin	DNase-accessible chromatin; TF footprinting	Nucleosome positioning, inaccessible chromatin	Location of target on DNA
Major advantage	Links labeling of accessible regions and next-generation sequencing library preparation, making preparation of library straightforward	Footprinting analysis	Method of choice for nucleosome positioning and quantitative nucleosome dynamics	Enables mapping of specific TF or histone modification in low cell numbers. Some histone modifications, like H3K27ac, can be used to look for active enhancers

from >12 samples in a single workday, and (ii) the assay requires little to no tuning or optimization of the reaction parameters for different cell types. We find that this protocol works without modification on a variety of cell lines and primary cell/tissue types. In contrast, DNase-seq and MNase-seq assays often require optimization of the enzyme concentration and the reaction time for the specific input material to avoid over- or underdigestion of chromatin, which influences the types of hypersensitive sites detected²⁸. ATAC-seq is a noncatalytic enzymatic reaction, where each Tn5 molecule can only perform a single transposition reaction, thus reducing the risk of chromatin overdigestion inherent in endonuclease assays. The Tn5 enzyme is commercially available with preloaded adapters, making assay setup simple and with little to no calibration required (see the ‘Tn5-to-cell ratio’ section of ‘Experimental design’ for more details). Since their inception, all of these methods have been adapted to the low input or single-cell scale^{64–67}, no longer limiting their applications to highly abundant cell types. ATAC-seq is a highly portable technique that can be applied to many different experimental scenarios and in any laboratory, regardless of previous experience with epigenomics, to answer general questions about the gene regulatory landscape. However, there are instances where ATAC-seq may not be the optimal assay of choice, notably for TF footprinting, for analyzing dynamics of nucleosome lability or when dealing with formalin-fixed paraffin-embedded (FFPE)-treated samples.

Comparison with previous ATAC-seq methods

The initial development of ATAC-seq³² made the study of gene regulation accessible to a much larger community of researchers. The assay was quickly applied to a variety of cell lines and primary cell types, enabling applications that were previously impossible. Despite this success, multiple shortcomings were noted in the original method. For example, since mitochondrial DNA is not chromatinized, the presence of lysed mitochondria in an ATAC-seq reaction can lead to a very high fraction of ATAC-seq reads mapping to mitochondrial DNA. In some applications, >97% of all reads mapped to mitochondrial DNA⁴⁴, making it difficult and costly to capture sufficient reads mapping to the desired nuclear genome. Additionally, low signal-to-background ratios in many cell types and contexts made application of ATAC-seq to certain experimental systems difficult or impossible. To overcome these challenges, researchers developed cell-type-specific optimizations, including the use of different detergents for cellular lysis such as digitonin³⁷ or Triton X-100⁶⁸, the addition of more transposase enzyme to the assay⁶⁹ or the use of clathrin inhibitors such as Pitstop 2⁷⁰. However, these optimizations were targeted for specific applications and do not necessarily extrapolate well to a wide variety of cell types, highlighting the need for more broadly applicable protocol optimizations. We have previously developed a generalizable and optimized version of the ATAC-seq method, called Omni-ATAC³⁶, that removed many of the cell- or context-specific problems limiting broad application of ATAC-seq.

Development of the Omni-ATAC protocol

The Omni-ATAC protocol improved upon the original ATAC-seq methodology by reducing reads mapping to mitochondrial DNA and increasing signal-to-background ratios across diverse cell lines, tissues and frozen samples³⁶. This was achieved through improvements to cell lysis, nuclei isolation and transposition. The optimizations in the Omni-ATAC protocol enable lysis of diverse cell types by including Tween-20 and digitonin in addition to Nonidet P40 (NP40). NP40 serves as the primary permeabilization reagent, allowing permeabilization of both plasma and nuclear membranes. Digitonin is a steroidal saponin that binds to cholesterol and related molecules that are highly enriched in the plasma membrane compared with intracellular membranes. Because of this, digitonin efficiently permeabilizes the plasma membrane but not the nuclear or mitochondrial membranes. Tween-20, which on its own is a weak permeabilization agent, serves primarily to prevent nonspecific binding of mitochondria to the nuclear membrane. In combination, these detergents provide broad-spectrum lysis of diverse cell types without over-lysis. The introduction of a post-lysis nuclei wash step with Tween-20 serves to remove mitochondria from the transposition reaction, thereby increasing library complexity and reducing sequencing costs. Omni-ATAC also includes Tween-20 and digitonin in the transposition reaction at 37 °C, ensuring adequate lysis of harder-to-lyse cell types. Finally, the addition of PBS to the transposition reaction simultaneously increases the signal-to-background ratio and the total number of unique fragments recovered. We believe that this effect is driven by the chaotropic nature of PBS, which may make the chromatin slightly more accessible for transposition. In combination, these optimizations enable the identification of more transposase-accessible peaks, improve data quality when profiling low cell numbers (>500 cells) and establish a single protocol that

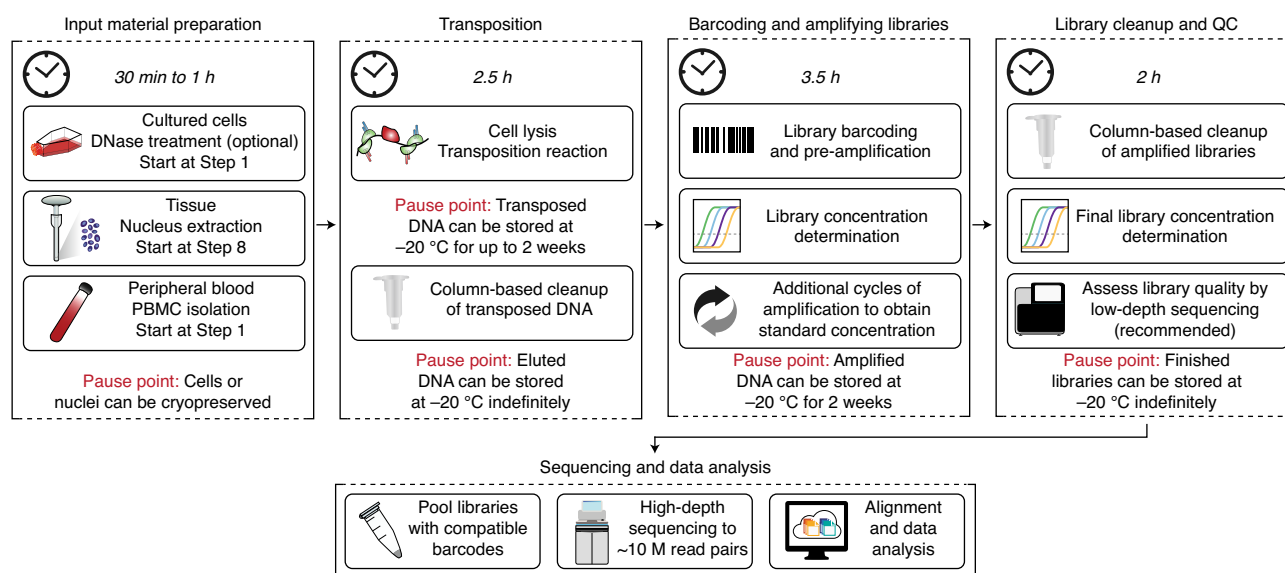


Fig. 2 | Schematic overview of ATAC-seq protocol. The primary protocol steps are shown in sequence, along with approximate timing and pause points.

can be applied to a wide variety of cell types, including many that could not be assayed using the original ATAC-seq protocol. The protocol described below represents an updated version of the Omni-ATAC protocol, incorporating a more modern polymerase, reducing the amount of primer used in amplification, changing how pre-amplification is performed and providing more explicit guidelines for library preparation and proper amplification of ATAC-seq libraries (Supplementary Fig. 1a–c).

Experimental design

Brief overview

ATAC-seq can be summarized in five stages: input material preparation, transposition, library preparation, sequencing and data analysis. This protocol focuses in detail on the transposition and library preparation stages, with guidance for how to obtain the proper input material and how finalized libraries should be sequenced and analyzed. We refer readers to several established bioinformatics pipelines and software tools for ATAC-seq analysis for this final stage^{71–73}. See the ‘Data analysis’ section for an overview of the analysis workflow. An overview of the experimental wet-lab protocol, with possible pause points, is shown in Fig. 2.

In brief, ATAC-seq library preparation involves (i) transposition into native chromatin, (ii) barcoding and amplification of transposed DNA and (iii) purification and quantification of libraries for sequencing. After sample collection (see ‘Input material preparation’), cells are lysed, releasing nuclei. The nuclear membrane is permeabilized in a buffer containing NP40, Tween-20 and digitonin. Transposition, using Tn5 enzymes complexed with double-stranded oligos containing PCR-compatible handles (Supplementary Fig. 2), is performed at 37°C , and the resulting DNA fragments are purified and barcoded by PCR. Finally, to ensure optimal amplification of all samples across conditions, the concentration of the pre-amplified library is quantified by qPCR and the necessary number of extra PCR cycles is determined. After this final amplification, ATAC-seq DNA is purified and ready for high-throughput sequencing. Below we detail some aspects of the experimental design to consider when setting up an ATAC-seq experiment.

Input material preparation

One of the most important considerations for a successful ATAC-seq experiment is the quality of the input material. The Omni-ATAC protocol is robust to a variety of mammalian cell and tissue types; however, some optimization of sample preparation may be necessary. Input material can be as low as 500 cells (or nuclei), although we find that optimal results are obtained with 50,000 cells and a comfortable lower limit for most researchers would be 5,000 cells. This protocol starts from fresh or cryopreserved whole cells or nuclei. Flash-frozen cell pellets or formaldehyde-crosslinked cells

Box 1 | DNase treatment of cultured cells

For optimal data quality, ensure that the cells are viable prior to performing ATAC-seq. We recommend viability >90%, preferably ~95%. To deplete highly viable cultures (85–95% viable) of free-floating DNA that originates from dead cells, cells can be treated with DNase, which may increase data quality. Cultured cells can be treated directly on the plate, adding DNase to a final concentration of 200 Kunitz units/mL directly to the cell culture medium followed by incubation at 37 °C for 30 min in a cell culture incubator. Be sure that the culture medium lacks EDTA, as DNase needs divalent cations for its activity. Remove 60,000 cells per ATAC-seq reaction (accounting for loss of cells with spinning and washing), and place in a 1.5 mL LoBind tube. Fill the tube with sterile PBS. Centrifuge the cells at RT for 5 min at 350g. Remove the supernatant, and resuspend the cells in 1 mL of sterile PBS. Centrifuge again at RT for 5 min at 350g. The resulting cell pellet can be used as input to the ATAC-seq protocol, starting at Step 9. If clear cell pellets are not obtained in PBS, try adding Tween-20 to a final concentration of 0.1% (wt/vol) or BSA to a final concentration of 0.5% (wt/vol).

generally give lower-quality libraries; thus, input cells should either be freshly cultured or stored frozen in a cryopreservative, such as BAM Banker (see ‘Reagents’). Other similar, serum-free cryopreservatives would also be suitable. We describe the most common input material types and provide general recommendations for handling. Note that the recommendations below are for mammalian cells, although ATAC-seq, with some modifications, has been applied to plants^{74,75}, whole *Caenorhabditis elegans*⁴⁵, *Drosophila* embryos⁷⁶, yeast⁶¹ and many other nonmammalian systems.

Cultured cells. If using freshly cultured mammalian cells, ensure that the cells are viable—dead cells often have large amounts of unchromatinized DNA, which is highly accessible to Tn5 and thus increases background noise. If the cells are >85% viable, pretreatment with DNase can remove any free-floating DNA (Box 1). For samples with lower viability, we recommend using a Ficoll gradient to remove dead cells (see ‘Reagents’ for product and associated protocol) or fluorescence-activated cell sorting (FACS) against a marker of dead cells such as 4',6-diamidino-2-phenylindole (DAPI), propidium iodide (PI) or Annexin V⁷⁷. Adherent cells can be lifted off the plate with trypsin or via other methods that are suitable for the cell type of interest while retaining viability. Suspension cells can be collected directly by centrifugation. Cells can also be cryopreserved and thawed at a later date for use in ATAC-seq. We recommend cryopreservation in aliquots of ~60,000 cells (accounting for some loss) in 100 µL of cryopreservative (see ‘Cryopreserved cells or nuclei’).

Whole blood. Peripheral blood mononuclear cells (PBMCs) should be isolated from whole blood, by either a Ficoll gradient or similar approaches. Because of their unique chromatin structure⁷⁸ and high levels of endogenous proteases, neutrophils present a challenge in ATAC-seq. We recommend ensuring the complete removal of neutrophils prior to performing ATAC-seq as it has been shown to greatly improve ATAC-seq data quality⁷⁹. If using cryopreserved PBMCs, we recommend pretreating the thawed cells with DNase prior to Ficoll or FACS as described in Box 1.

Tissues. For complex tissues, including flash-frozen tissue, we recommend extracting nuclei from the tissue prior to starting the transposition reaction, as direct tissue lysis with the ATAC-seq Lysis Buffer is not efficient. We previously established a nuclei isolation protocol that performs well for a variety of human and nonhuman tissues^{36,80}, and we include a detailed version of this protocol in Supplementary Protocol 1. When collection of samples makes the use of fresh tissue impossible, tissues should be flash-frozen or stored in cryopreservative, prior to nucleus extraction. A picture of nucleus input material is given in Supplementary Fig. 3a. If using Supplementary Protocol 1 for nucleus extraction, the resulting high-quality nuclei can be pelleted and treated as the input material for transposition, beginning at Step 8 of the main Procedure. Because these nuclei were isolated in the presence of NP40, there is no need for additional lysis with NP40. Nuclei can also be cryopreserved. We recommend cryopreservation in aliquots of ~60,000 nuclei (accounting for some loss) in 100 µL of cryopreservative, which can be used as described below (see ‘Cryopreserved cells or nuclei’ and Supplementary Protocol 1). We note that many nuclei isolation protocols have been published and certain tissue types require specialized procedures for nuclei isolation. Table 2 provides a non-exhaustive list of references for a variety of nuclei isolation protocols that could be considered as alternatives.

Cryopreserved cells or nuclei. If cryopreserving cells, it is important to ensure that the cells are viable prior to cryopreservation. Generally speaking, it is the cell state at the time of cryopreservation that

Table 2 | Published nucleus extraction protocols

Input material	Method of isolation and purification	Reference	Notes
Cells, tissues	Douncing; iodixanol gradient centrifugation	Supplementary Protocol 1	Validated for use with ATAC-seq for a wide range of tissues, especially those with low extracellular matrix content (ECM). Protocol is derived from several methods ^{36,38,141}
Tissues with high ECM or vasculature content	Incubation in lysis buffer and gentle trituration with 5 mL serological pipette	Yang et al. ¹⁴²	Good for difficult-to-dissociate tissues, validated for choroid plexus scRNA-seq; can capture vascular cells often lost in other isolation methods
Tissues; enteric neurons	Chopping with scissors on ice + FACS	Drokhlyansky et al. ¹⁴³	Validated for mouse enteric neurons in scRNA-seq; retains a larger percentage of endoplasmic-reticulum-associated RNAs
Tagged nuclei (INTACT)	Expression of nuclear targeting fusion protein; bead purification	Deal and Henikoff ^{75,144}	Developed originally for plants, but has been applied to many systems
Fibrous tissues (PAN-INTACT); validated for heart and kidney PBMCs, cell lines	Douncing; filtration; sucrose cushion centrifugation	Bhattacharyya et al. ¹⁴⁵	Show that changing the concentrations of the sucrose gradient can enrich for different cell types
	Lysis and centrifugation	10x Demonstrated Protocols ¹⁴⁶	Validated for use with the 10x genomics scATAC-seq assay
Plant cells	Sucrose sedimentation; streptavidin bead separation of tagged nuclei	Bajic et al. ⁷⁴	Has been used in several plant types, including <i>Arabidopsis</i> , tomato and rice. INTACT system can be applied to other nonplant systems, supposing that they have a nuclear tag, e.g., the use of the SUN1-GFP tag in mammalian cells ¹⁴⁷
Yeast	Spheroplast centrifugation or homogenization or manual rupture of frozen cells	Kiseleva et al. ¹⁴⁸	Several options given; similar to other published protocols ^{61,149}
Tissues, especially brain tissue	Mild fixation, douncing, sucrose centrifugation and FACS	Nott et al. ¹⁵⁰	Validated for brain tissues; FACS sorting to obtain cell type specific nuclei (e.g., NeuN, PU.1 or OLIG2 staining). Great for selecting a target cell of interest in the brain. Similar strategy used in other papers, without fixation ^{151,152}
<i>Drosophila</i> embryos	Freeze, lysis buffer and crush	Haines ¹⁵³	Based on protocols similar to the one described in this publication, but with embryo-specific lysis steps
<i>Caenorhabditis elegans</i>	Based on the INTACT methods. Biotin labeling of nuclear envelope	Steiner et al. ¹⁵⁴	Especially for muscle tissue. Also used for <i>Drosophila</i> . Another method for <i>C. elegans</i> germ cells ¹⁵⁵

will be assessed by ATAC-seq. Upon thaw, some plasma membrane lysis will occur, but the nuclear membrane and chromatin will be preserved if cells or nuclei are maintained at 4 °C. We therefore do not worry about cells ‘dying’ during the thaw. To thaw aliquots of ~60,000 cryopreserved cells/nuclei, we recommend adding 1 mL of cold ATAC-seq Wash Buffer directly to the 100 µL of frozen cryopreservative. Incubate on ice until the cryopreservative has thawed, and proceed to Step 3 for cells or Step 8 for nuclei. We use one aliquot of ~60,000 cells or nuclei for a single ATAC-seq reaction and assume that this results in ~50,000 cells or nuclei, accounting for the extra handling and centrifugation steps involved in freezing and thawing. To thaw cells or nuclei that have been cryopreserved in aliquots of more than ~60,000 cells or nuclei, we recommend thawing the cryopreservative on ice, counting cell density if needed, removing the desired number of cells or nuclei, washing them in 1 mL of cold ATAC-seq Wash Buffer and proceeding to Step 3 for cells or Step 8 for nuclei. We find that nuclei stored in cryopreservative can be freeze-thawed multiple times with only very moderate effects on downstream data quality.

Input types that do not work for ATAC-seq. As of the writing of this manuscript, application of ATAC-seq to FFPE tissue is not widely established, though published methods have been developed⁸¹. However, even when using methods that have been tailored to FFPE, ATAC-seq library quality from FFPE tissues remains much lower than other input material types. Alternative non-ATAC-seq methods for profiling chromatin accessibility in FFPE tissues have also been developed⁸². Nevertheless, we recommend avoiding the use of flash-frozen cell pellets, formaldehyde-crosslinked

cells or FFPE tissues, if possible, as these methods yield lower-quality ATAC-seq data and are not well suited as input to the Omni-ATAC protocol.

Biological versus technical replicates

We generally advise the use of biological rather than technical replicates, when resources are constraining. When the number of available biological replicates is constraining, it may be advantageous to perform two or three technical replicates to ensure reproducible peak capture. Technical replicates, i.e., those derived from the same biological material, should be combined at the peak merging step of analysis (see ‘Peak calling, merging and counting and annotation’). Additionally, it is important to note that many differential accessibility tools will expect at least two biological samples per condition for accurate modeling and statistics.

Tn5-to-cell ratio

Recent work has shown that the ratio of Tn5 to cells has an effect on data quality⁸³. To standardize the transposition reaction, we suggest keeping the ratio of cells to Tn5 constant across all samples by standardizing the number of cells or nuclei in the ATAC-seq reactions. This helps to ensure similar TSS enrichment scores across different input amounts (Supplementary Fig. 4). The Tn5 ratio used in the main protocol assumes the use of 50,000 cells in the reaction. Refer to Supplementary Note 2 if using more or fewer than 50,000 cells.

Quality control of ATAC-seq libraries

We strongly recommend determining the quality of final ATAC-seq libraries through low-depth sequencing (50,000–100,000 read pairs per sample). The success of ATAC-seq library generation is determined by four key factors: (i) the enrichment of transposase insertions in regions of known chromatin accessibility (signal-to-background ratio), (ii) the total number of unique fragments (library complexity), (iii) the ratio of sequencing reads mapping to the nuclear genome (desired) versus those mapping to the mitochondrial genome (undesired) and (iv) the fragment size distribution.

Of these, the signal-to-background ratio is the single-most important quality control (QC) metric for ATAC-seq. Other epigenomic assays such as ChIP-seq have often used the fraction of reads in peaks to determine signal-to-background ratios. This requires a priori knowledge of the location of peak regions or sufficient sequencing depth to call peak regions on a per-sample basis, the latter of which is flawed as it yields incomparable results across samples of differing sequencing depths. Instead, we summarize this signal-to-background ratio using a TSS enrichment score, which quantifies the relative enrichment of signal in regions surrounding TSSs (Fig. 3a–c) because these promoter regions are known to be consistently enriched in accessible chromatin³². Accurate TSS enrichment scores for an ATAC-seq library can be obtained from as few as 50,000 read pairs. TSS enrichment scores were introduced in the original description of ATAC-seq and have been adopted as the standard by multiple ATAC-seq pipelines^{72,73,84} (see ‘Data analysis’). It is important to note that the numeric value of the TSS enrichment score is dependent on the set of TSSs used in the calculation, which means that scores may not be directly comparable across different tools or species and each pipeline will provide recommendations for passable scores.

Low-depth sequencing can also be used to assess the ratio of sequencing reads mapping to the nuclear and mitochondrial genomes and the fragment size distribution. All of these metrics are standard outputs of the ATAC-seq pipelines recommended below. If low-depth sequencing is not readily available, the shape of a Bioanalyzer or TapeStation trace can be used to determine the fragment size distribution. While we do not recommend using this as a proxy for quality, it is often true that the library is of passable quality if it shows some nucleosomal periodicity (Fig. 3a). However, we would caution that some good-quality libraries do not have the clear nucleosomal periodicity banding, and this does not necessarily imply that they have been under transposed. For example, two libraries with similar TSS enrichment scores (8.3 and 8.8, respectively) have different Bioanalyzer traces, and the relative absence of the nucleosomal banding alone does not imply a library has a lower TSS enrichment score (Fig. 3d,e). Compared with the original ATAC-seq method, the Omni-ATAC protocol generates ATAC-seq libraries that are biased toward higher-molecular-weight fragments, likely due to the inclusion of PBS in the reaction. Occasionally, the ratio of higher-molecular-weight fragments to lower-molecular-weight fragments is sufficiently high to obscure the nucleosomal periodicity of the lower-molecular-weight fragments. However, the presence of these high-molecular-weight

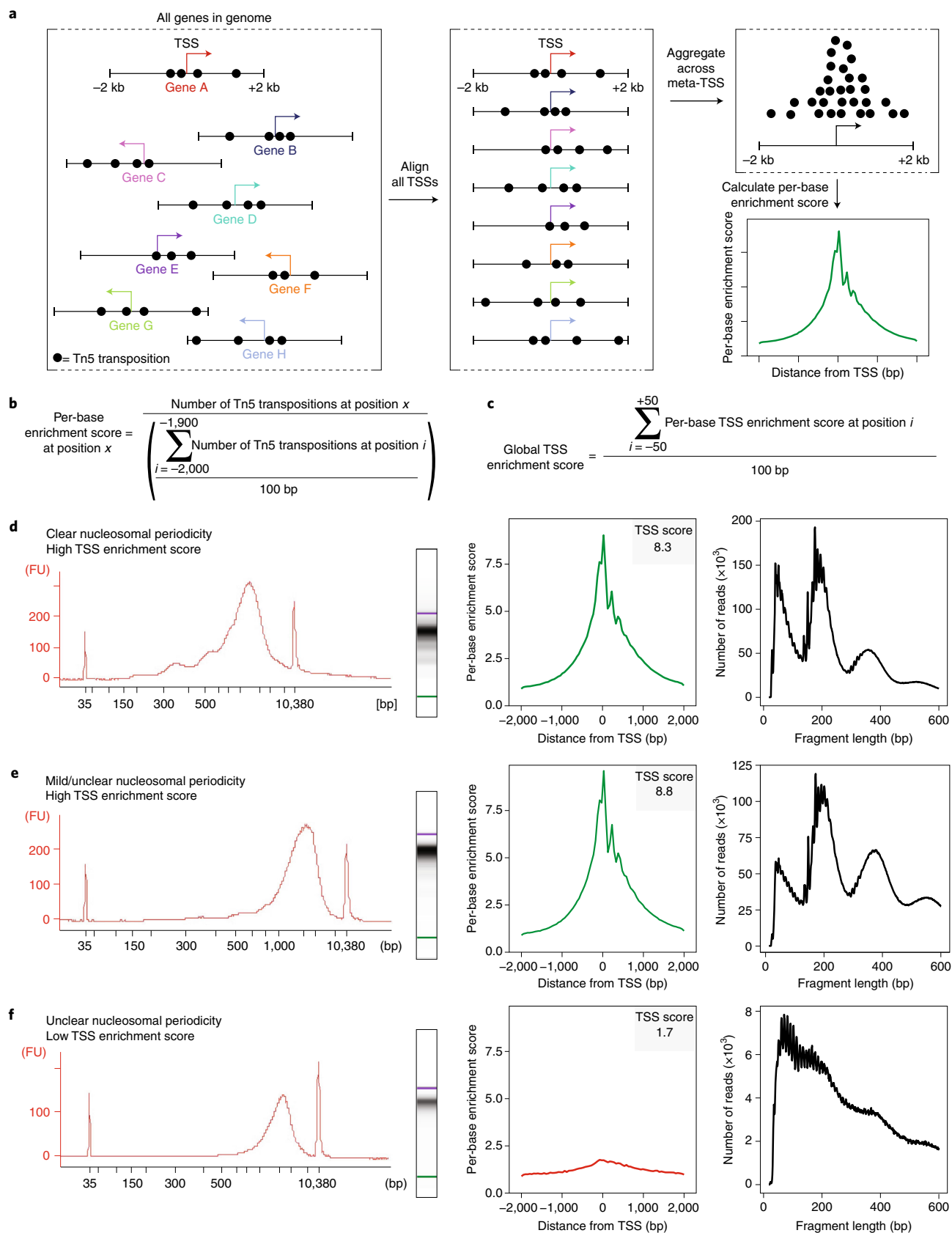


Fig. 3 | Assessing ATAC-seq library quality. **a**, A schematic showing transposition events at a set of genes. The TSS is depicted as an arrow, with the direction of the arrow indicating whether the gene is present on the plus or minus strand. Each transposition event is shown as a black circle. Each gene has been clipped to include the region ± 2 kb from the annotated TSS, and this region will be used in the calculation of the TSS enrichment score. Next, the TSS location and direction for every gene in the genome are aligned and the reads are aggregated across all of these sites into a 'meta-TSS'. This then allows for the calculation of a per-base enrichment score via the equation in **b**, which can be plotted along the ± 2 kb of the meta-TSS, as shown in the TSS enrichment score graph in the bottom right. **b**, The equation, implemented by the PEPATAC pipeline, for calculating the per-base enrichment score. **c**, The equation, implemented by the PEPTATAC pipeline, for calculating the global TSS enrichment score for the whole library. **d–f**, From left to right: Bioanalyzer electropherogram trace and digitally rendered gel (left), TSS enrichment plot (middle) and fragment size distribution (right). Libraries shown are a successful ATAC-seq library with a high TSS enrichment score (8.3) and clear nucleosomal periodicity in the Bioanalyzer trace (**d**), a successful ATAC-seq library with a high TSS enrichment score (8.8) but minimal observed nucleosomal periodicity in the Bioanalyzer trace (**e**), and an unsuccessful ATAC-seq library with a low TSS enrichment score (1.7) and no clear nucleosomal periodicity in the Bioanalyzer trace (**f**). Note that all three libraries (**d–f**) have abundant high-molecular-weight fragments ($>2,000$ bp) on the Bioanalyzer trace (left), but these fragments do not cluster on the sequencer and thus are not represented in the fragment size distributions derived from the corresponding sequencing data (right). In **e**, despite not being easily observed on the Bioanalyzer trace, the characteristic fragment size distribution is observed using low-depth sequencing. In **f**, the TSS enrichment score and fragment size distribution plots on the right indicate a poor signal-to-background ratio and low sample quality, respectively, marking this library as a poor candidate for high-depth sequencing.

fragments does not imply that the library is undertransposed. These larger fragments do not cluster well on Illumina sequencers and can be safely ignored (Fig. 3d–f).

Sequencing of final ATAC-seq libraries

ATAC-seq data should always be sequenced using paired-end sequencing because each end of an ATAC-seq fragment corresponds to a unique Tn5 transposition event (Fig. 1b). Therefore, using single-end sequencing effectively ignores 50% of the data. For high-depth sequencing of final libraries we recommend targeting 10 million read pairs (10 million clusters passing filter). This is sufficient for most routine analyses including differential accessibility analysis and motif enrichment analysis. However, when genotyping of the ATAC-seq data will be performed, longer reads can be used to capture more DNA sequence. Detailed guidelines are provided in Table 3. Because both ends of the ATAC-seq fragments have 8 bp barcodes, two 8 bp indexing reads should be performed. We provide adapter and barcode sequences in Supplementary Table 2. The minimum read length will be dependent on the particular application of the ATAC-seq library; for most standard applications, 36 base pairs is sufficient and provides equivalent mapping efficiency to longer reads from a variety of input materials (Supplementary Table 1). Mapping rates and TSS enrichment scores for a variety of different read lengths are provided in Supplementary Table 1.

Data analysis

After sequencing, we recommend using publicly available pipelines to perform alignment and downstream analysis (see 'Anticipated results'). PEPATAC⁷³ is a highly portable and user-friendly pipeline that provides multiple quality metrics, including the TSS enrichment score. ENCODE⁷² and nf-core⁸⁴ also provide similarly robust pipelines for ATAC-seq analysis. Any of these pipelines can be used for analysis of low-depth QC data or high-depth sequencing data. Regardless of the pipeline, ATAC-seq data analysis follows these general steps: pre-alignment sequence QC, trimming of adapter sequences, alignment to the reference genome of interest, removal of reads mapping to mitochondrial DNA and PCR duplicates, and peak calling. This is often followed by more application-specific analyses such as differential accessibility testing, motif enrichment analysis and integration with other data types such as RNA-seq, ChIP-seq or chromatin conformation capture.

Single-cell ATAC-seq

The Omni-ATAC protocol described here was designed for bulk ATAC-seq. However, the commercially available scATAC-seq products from 10x Genomics utilize similar lysis and transposition reaction conditions. For researchers interested in performing scATAC-seq, we recommend following the manufacturer instructions for standard cell types. For frozen tissues, we provide recommendations for how to isolate nuclei to use as input to the 10x Genomics scATAC-seq kit in our nuclei isolation protocol (Supplementary Protocol 1)⁸⁰.

Necessary expertise

Basic molecular biology skills are necessary to perform ATAC-seq. Sequencing of the final ATAC-seq libraries requires the use of high-throughput sequencing platforms usually found in a genomics core

Table 3 | Guidelines for sequencing parameters

Application	Insight gained	Minimum read length ^a	Index length ^b	Paired- or single-end	Read pairs per sample
Gene regulatory landscape profiling	Peaks, differential peaks between samples, motif analysis of peaks	36 bp	8	Paired	10 million
Genotyping	Gene regulatory landscape + genotype of sample; useful for patient samples and to determine whether sequence variants affect a peak	100 bp	8	Paired	10 million
Footprinting analysis	Footprinting of different TFs to determine binding sequence at base-pair resolution	36 bp	8	Paired	200 million
Nucleosome occupancy	Location of nucleosomes along DNA	36 bp	8	Paired	60 million

^aLonger read lengths can be used but may not provide additional gains. We find that 36 bp is sufficient to map transposition sites. Given that most ATAC-seq fragments are <200 bp in length, sequencing reads longer than 100 bp do not provide much additional data. ^bIndex length assuming the barcodes listed in Supplementary Table 2 are used.

facility. For data analysis, access to a computational server or high-performance computing environment may be necessary, as processing data on this scale can be computationally intensive for most standard computers. Basic familiarity with command-line tools and bioinformatics is necessary for pipeline implementation.

Limitations

As discussed above, ATAC-seq gives a broad picture of the epigenetic landscape of a sample; however, it cannot provide specific details about the exact mechanism of action (i.e., which chromatin marks or TFs are present on the region of interest). As such, ATAC-seq data often benefit from correlation with other data types for interpretation. Additionally, ATAC-seq is not very well suited for TF footprinting of an individual genomic locus, which is used to identify the precise base-pair resolution sequence bound by a TF. Such goals would require very high-depth sequencing data and may be better served by other techniques such as DNase-seq^{27–29}. Similar to DNase-seq^{28,85}, ATAC-seq is known to have sequence biases, based on the insertion preferences of Tn5³². Additionally, the protocol detailed here obtains ATAC-seq data from bulk cell populations, and thus presupposes that there is limited intercellular heterogeneity or that this heterogeneity is not of interest. If the input material is highly heterogeneous, for example, a complex tissue, then the resultant ATAC-seq profile will represent the average signal of all the cells and cell types, thus lacking signatures of rare cell types (<20% of the total cells)⁵³. To adequately capture this heterogeneity, it is either necessary to perform single-cell/single-nucleus ATAC-seq (snATAC-seq)^{86,87} or to FACS-purify the desired population of cells and perform bulk ATAC-seq on that purified population. Finally, it is important to note that ATAC-seq can only help give regulatory insights for phenomena that are regulated at the transcriptional level and will inherently miss any post-transcriptional changes such as RNA decay or modification, export or translation.

Materials

Biological materials

- Cell line(s) or input material of interest

Reagents

- PBS (Thermo Fisher Scientific, cat. no. 10010049)
- 1 M Tris-HCl pH 7.5 (Thermo Fisher Scientific, cat. no. 15-567-027)
- 5 M NaCl (Corning, cat. no. 46-032-CV)
- 1 M MgCl₂ (Invitrogen, cat. no. AM9530G)
- UltraPure DNase/RNase-free distilled water (Thermo Fisher Scientific, cat. no. 10977015)
- 2% (wt/vol) digitonin (Promega, cat. No. G9441) **! CAUTION** Digitonin solutions can cause skin and eye irritation. Handle using appropriate protective gloves. **▲ CRITICAL** Digitonin solutions can be difficult to make, and the dry reagent is acutely toxic. For best results, purchase 2% digitonin and dilute as described in 'Reagent setup'.

- 10% (wt/vol) Tween-20 (Millipore-Sigma, cat. no. 11332465001) **▲ CRITICAL** 100% Tween-20 can be difficult to dilute accurately. For best results, purchase 10% Tween-20.
- 10% (wt/vol) NP40 Substitute (Millipore-Sigma, cat. no. 11332473001) **! CAUTION** NP40 can cause skin and eye irritation. Handle using appropriate protective gloves. **▲ CRITICAL** 100% NP40 Substitute can be difficult to dilute accurately. For best results, purchase 10% NP40.
- Sequencing Adapters (IDT, see Supplementary Table 2)
- NEBNext Ultra II Q5 2× Master Mix (New England Biolabs, cat. no. M0544S) **▲ CRITICAL** We have optimized the ATAC-seq library preparation using the NEBNext Ultra II Q5 2× Master Mix. Other PCR master mixes may be substituted, but optimization may be necessary.
- Tagment DNA (TD) TDE1 Enzyme and Buffer Kit (Illumina, cat. no. 20034197 (48 reactions) or 20034198 (96 reactions)) **▲ CRITICAL** We have optimized the volume of Tn5 transposase per ATAC-seq reaction based on products purchased from Illumina. As the manufacturer does not provide an enzyme concentration, it is difficult to translate this to other Tn5 preparations. Thus, although other suppliers can be used, this will require additional optimization. It is also possible to produce and purify Tn5 enzyme in the laboratory (Supplementary Note 3).
- DNA Clean and Concentrator-5 Kit (Zymo Research, cat. no. D4014)
- NEBNext Library Quant Kit (New England Biolabs, cat. no. E7630)
- BAM Banker Cryopreservative (Fisher Scientific, cat. no. NC9582225)
- DNase (optional; Worthington, cat. no. LS002007)
- Ficoll Paque Plus (optional; Cytiva Life Sciences cat. no. 17144002)
- Hanks' Balanced Salt Solution, with calcium and magnesium, no phenol red (for DNase, optional; Thermo Fisher Scientific, cat. no. 14025092)
- Qubit 1× dsDNA HS Assay Kit (optional; Thermo Fisher Scientific, cat. no. Q33231)

Equipment

- 2 µL, 20 µL, 200 µL and 1,000 µL single-channel pipettes (Rainin, cat. nos. 17014393, 17014392, 17014391 and 17014382)
- 20 µL, 200 µL and 1,000 µL filter tips (Rainin, cat. nos. 17014961, 17014963 and 17014967)
- 5 mL, 10 mL, 25 mL and 50 mL serological pipettes (Thermo Fisher Scientific, cat. nos. 07-200-573, 07-200-574, 07-200-575 and 07-200-576)
- Pipet-Aid XP (Drummond, cat. no. 4-000-101)
- 1.5 mL DNA LoBind tubes (Eppendorf, cat. no. 4043-1021) **▲ CRITICAL** LoBind tubes should be used to maximize sample recovery of nucleic acids as they reduce sample-to-surface binding.
- Eight-strip DNase-free PCR tubes (USA Scientific, cat. no. 1402-4700)
- Vortex mixer (VWR, cat. no. 97043-562)
- CFX Opus 384 Real-Time PCR System (Bio-Rad, cat. no. 12011452)
- C1000 Touch thermal cycler (Bio-Rad, cat. no. 1851148)
- MyFuge 12 tabletop centrifuge (Benchmark Scientific, cat. no. C1012)
- Microcentrifuge (Eppendorf, cat. no. 2231000768)
- Refrigerated microcentrifuge (Eppendorf, cat. no. 5404000413)
- 15 mL and 50 mL conical tubes (Thermo Fisher Scientific, cat. nos. 05-538-53F and 05-538-55A)
- 1.5 mL microcentrifuge tubes (Thermo Fisher Scientific, cat. no. 14-222-155)
- 384-well plates (Bio-Rad, cat. no. HSP3905)
- Microseal 'B' PCR plate sealing film (Bio-Rad, cat. no. MSB1001)
- MPS 1000 plate centrifuge (Labnet International, cat. no. C1000)
- Thermomixer (Eppendorf, cat. no. 2231000680)
- 0.22 µm 250 mL cellulose nitrate sterilizing filter (Corning, cat. no. 430756)
- 0.22 µm 50 mL PVDF Steriflip-GV Sterile Centrifuge Tube Top Filter Unit (for DNase, optional; Millipore-Sigma, cat. no. SE1M179M6)
- Qubit 4 fluorometer (optional; Thermo Fisher Scientific, cat. no. Q33238)
- Vacuum manifold (optional; QIAGEN, cat. no. 19413)
- VacConnectors sterile adapters for vacuum manifold (optional; QIAGEN, cat. no. 19407)
- High-performance computing environment or similar

Software

- Several options: see 'Anticipated results' and Table 4
- Iterative overlap peak merging script⁸⁸: see 'Code availability'

Table 4 | Data analysis pipelines available for ATAC-seq

Step/process	ENCODE ATAC-seq	PEPATAC	nf-core atacseq
Version used in comparison	v1.10.0	v0.10.0	v1.2.1
Environment	Cromwell/caper	Pypiper	Nextflow
Trimming, alignment and deduplication	Cutadapt ¹⁵⁶ bowtie2 ¹⁶⁰ Picard ¹⁶²	TRIMMOMATIC ¹⁵⁷ or skewer ¹⁵⁸ bowtie2 ¹⁶⁰ or BWA ¹⁶¹ samblaster ¹⁶³ or Picard	TrimGalore! ¹⁵⁹ BWA Picard
Tn5 offset correction	Yes	Yes	No
Mitochondrial filter	Yes	Yes	Yes
Peak calling method	MACS2 ⁹²	MACS2 (default), F-seq ¹⁶⁴ or Genrich ⁹³	MACS2
Peak merging method	Based on the irreproducible discovery rate for replicates; does not merge for a whole set of samples	Fixed width, iterative overlap	Raw peak overlap using bedtools ¹⁰⁸ merge
Output provided	BAM files, bigwig files (one representing fold enrichment over expected background and the other representing statistical significance), BED file of peaks for each file and for the merged peak set	QC plots including alignment scoring, TSS scores and library complexity, BED peaks and counts, BAM files, bigwig files (nucleotide resolution and smoothed)	QC html report, BAM files, normalized bigwig files, BED peaks, annotation of peaks (HOMER), merged peak set, differential accessibility (DESeq2), Integrative Genomics Viewer output
Code repository	https://github.com/ENCODE-DCC/atac-seq-pipeline	https://github.com/databio/pepatac	https://github.com/nf-core/atacseq

Reagent Setup

ATAC-seq resuspension buffer (ATAC-RSB)

For 100 mL, combine 1 mL of 1 M Tris-HCl pH 7.5, 200 µL of 5 M NaCl, 300 µL of 1 M MgCl₂ and 98.5 mL of UltraPure distilled water. Filter sterilize using a 0.22 µm filter. The final composition of ATAC-RSB is 10 mM Tris-HCl pH 7.5, 10 mM NaCl and 3 mM MgCl₂ in water. Store at 4 °C for up to 6 months.

Digitonin

The recommended digitonin from Promega is supplied at a 2% (wt/vol) concentration in DMSO. Dilute this digitonin 1:1 with water to make a 1% (wt/vol) (100×) stock solution. This DMSO–water mixture will no longer freeze at –20 °C. Store at –20 °C for up to 6 months.

Tn5 transposase enzyme and 2× TD buffer

The Tn5 transposase enzyme precomplexed with adapters (TDE1) and the 2× TD buffer should be purchased from Illumina (see ‘Materials’). We strongly recommend purchasing these reagents as they are the core components of the reaction. However, protocols for producing, purifying and complexing the Tn5 transposase enzyme have been published⁸⁹ and are discussed briefly in Supplementary Note 3.

NEBNext Library Quantification Kit

Following the manufacturer’s instructions, add the primer to the enzyme mix. If using ROX reference dye (carboxyrhodamine) for normalization, add ROX as directed by the manufacturer.

Barcoding adapter resuspension

If ordering the barcoding adapters (Supplementary Table 2), reconstitute the oligonucleotides to 100 µM stock solutions in water and make 5 µM working stock solutions with water. Both solutions can be stored at –20 °C for 5 years or more.

DNase resuspension

If cells will be pretreated with DNase (optional), lyophilized DNase should be resuspended in Hanks’ Balanced Salt Solution to 20,000 Kunitz units/mL for a 100× stock solution. Filter sterilize this

solution using a 0.22 μm PVDF membrane. Aliquot in 100 μL to 1 mL volumes (according to the application), and store at $-20\text{ }^{\circ}\text{C}$ for up to 2 years. Once thawed, an aliquot can be stored at $4\text{ }^{\circ}\text{C}$ for up to 1 month but should not be refrozen.

Procedure

Transposition ● Timing 2.5 h for ~12 samples

- 1 Prepare samples according to the 'Input material preparation' section above. Here we assume that 50,000 freshly cultured cells have been collected in a 1.5 mL LoBind tube for each sample. For the purposes of this protocol, we refer to the starting material as cells, but the same protocol applies to nuclei. See Box 1 for optional DNase treatment.

▲ CRITICAL STEP We recommend piloting the cell or nuclei isolation ahead of time to make sure that visible pellets can be obtained after centrifugation and to practice supernatant aspiration in Step 4. If using the recommended 50,000 cells or nuclei, a small pellet should be visible.

? TROUBLESHOOTING

- 2 Prior to starting, make the ATAC-seq Lysis Buffer and ATAC-seq Wash Buffer and keep them on ice. Be sure to use freshly made ATAC-seq Lysis Buffer and ATAC-seq Wash Buffer each time.
- 3 Pellet 50,000 viable cells at 500g for 5 min at $4\text{ }^{\circ}\text{C}$ in a fixed-angle microcentrifuge. Orient the tubes in a consistent fashion so that the pellet will be visible on the outer surface of the tube toward the bottom.
- 4 Aspirate all the supernatant using two pipetting steps. First, aspirate down to 100 μL with a p1000 pipette. Then, remove the final 100 μL with a p200 pipette.
▲ CRITICAL STEP Make sure to avoid the visible cell pellet when pipetting. Optimal removal of supernatant and minimal disruption of the cell pellet is attained when the removal of the final 100 μL is performed in a consistent and fluid motion without starting and stopping.
- 5 Resuspend the cell pellet in 50 μL of ATAC-seq Lysis Buffer by pipetting up and down three times. ATAC-seq Lysis Buffer should be made fresh each time and mixed thoroughly prior to use.

ATAC-seq Lysis Buffer

Reagent	Volume per sample (μL)	Final concentration
Cold ATAC-RSB	48.5	
10% (wt/vol) NP40	0.5	0.1% (wt/vol)
10% (wt/vol) Tween-20	0.5	0.1% (wt/vol)
1% (wt/vol) digitonin	0.5	0.01% (wt/vol)
Total volume	50	

- 6 Incubate on ice for 3 min. If lysing multiple samples, make sure that all samples are lysed for the same total amount of time by proceeding to Step 7 after 3 min. An example image of lysed GM12878 cells is given in Supplementary Fig. 3b.
- 7 Add 1 mL of ATAC-seq Wash Buffer to dilute the lysis reagents. Invert the tube five times to mix. ATAC-seq Wash Buffer should be made fresh each time and mixed thoroughly prior to use.

ATAC-seq Wash Buffer

Reagent	Volume per sample (μL)	Final concentration
Cold ATAC-RSB	990	
10% (wt/vol) Tween-20	10	0.1% (wt/vol)
Total volume	1,000	

- 8 Pellet nuclei at 500g for 10 min at $4\text{ }^{\circ}\text{C}$ in a fixed-angle microcentrifuge. Orient the tubes in a consistent fashion so that the pellet will end up in the same location. If using nuclei isolated from frozen tissues as described in 'Input material preparation' above, start the ATAC-seq protocol here using 50,000 nuclei resuspended in 1 mL of ATAC-seq Wash Buffer.

- 9 Aspirate all supernatant using two pipetting steps as above. First, aspirate down to 100 μL with a p1000 pipette. Then, remove the final 100 μL with a p200 pipette.
▲ CRITICAL STEP Make sure to avoid the visible cell pellet when pipetting. Optimal removal of supernatant and minimal disruption of the cell pellet is attained when the removal of the final 100 μL is performed in a consistent and fluid motion without starting and stopping.
- 10 Resuspend the cell pellet in 50 μL of Transposition Mix by pipetting up and down six times. Transposition Mix should be made fresh each time and mixed thoroughly prior to use.

Transposition mix

Reagent	Volume per sample (μL)	Final concentration
2 \times TD buffer	25	1 \times
PBS	16.5	
UltraPure distilled H_2O	5	
1% (wt/vol) digitonin	0.5	0.01% (wt/vol)
10% (wt/vol) Tween-20	0.5	0.1% (wt/vol)
TDE1 TD enzyme (Tn5 transposase)	2.5	
Total volume	50	

- 11 Incubate reaction at 37 $^{\circ}\text{C}$ for 30 min in a thermomixer with 1,000 rpm mixing.
- 12 Remove the tubes from the thermomixer, and immediately terminate the transposition reaction by adding 250 μL (five volumes) of DNA Binding Buffer from the DNA Clean and Concentrator-5 Kit and mix well by pipetting or inversion.
▲ CRITICAL STEP We strongly recommend the use of different kits/reagents for the cleanup of pre- and post-amplification products so as to not contaminate post-amplification products into pre-amplification samples.
- 13 Pulse centrifuge to collect solution in the bottom of the tube.
■ PAUSE POINT This solution can be stored at -20°C for up to 2 weeks. Allow this mixture to warm back to room temperature (RT, 22 $^{\circ}\text{C}$) and mix thoroughly before proceeding.
- 14 Clean up the transposition reaction using the DNA Clean and Concentrator-5 Kit. If using the vacuum manifold, we recommend using sterile single-use VacConnectors to prevent cross-contamination. Transfer each sample, mixed with the DNA Binding Buffer, to a Zymo-Spin Column in a collection tube. Centrifuge at RT for 30 s at 10,000g, and discard the flowthrough.
- 15 Add 200 μL of DNA Wash Buffer to the column, and centrifuge at RT for 30 s at 10,000g.
- 16 Repeat this wash for a total of two wash steps.
- 17 Perform a final 'dry spin' after the second wash step to remove any traces of residual wash buffer from the column membrane. To do this, remove any flowthrough from the collection tube and centrifuge the column and collection tube at RT for 1 min at $>13,000g$.
- 18 Transfer the column to a clean prelabeled 1.5 mL LoBind tube. Pipette 21 μL of Elution Buffer directly onto the column membrane, and wait for 1 min.
- 19 Centrifuge the column at RT for 1 min at 13,000g to elute the DNA. This elution volume typically results in 20 μL of product.
■ PAUSE POINT This solution can be stored at -20°C for as long as necessary.

Barcoding of transposed fragments ● Timing 30 min

- 20 Assign each sample in the study to a unique combination of the Adapter 1 (Ad1) and Adapter 2 (Ad2) sequences found in Supplementary Table 2, and record this information. These combinations will be used to assign each read pair to its appropriate sample designation.
▲ CRITICAL STEP The adapters contain the dual-indexing sample-specific barcodes; therefore, each individual sample should receive a unique combination of Ad1 and Ad2. More explicitly, it is ok if two different samples share the same Ad1 as long as they can be differentiated by their Ad2, or vice versa. Samples with identical combinations of Ad1 and Ad2 cannot be sequenced together as the index reads derived from each sample will be indistinguishable.
- 21 Transfer each cleaned-up transposed DNA sample to a 200 μL PCR tube.
- 22 Add 25 μL of NEBNext Ultra II Q5 2 \times Master Mix to each tube.

- 23 Add 2.5 μL of the corresponding Ad1 to each sample.
- 24 Add 2.5 μL of the corresponding Ad2 to each sample.
- 25 Cap tubes, vortex and spin down to collect all liquid at the bottom of the tube. When completed, each reaction should contain the following:

Barcoding PCR reaction (per sample)

Reagent	Volume (μL)	Final concentration
Transposed sample	20	
NEBNext Ultra II Q5 2 \times Master Mix	25	1 \times
5 μM Adapter Ad1 ^a	2.5	0.25 μM
5 μM Adapter Ad2 ^a	2.5	0.25 μM
Total volume	50	

^aSee Supplementary Table 2 for adapter sequences.

- 26 Run the barcoding PCR reactions according to the following cycling conditions:

Cycle no.	Denature	Anneal	Extend
1			72 $^{\circ}\text{C}$, 5 min
2	98 $^{\circ}\text{C}$, 30 s		
3–5 (3 cycles)	98 $^{\circ}\text{C}$, 10 s	65 $^{\circ}\text{C}$, 30 s	65 $^{\circ}\text{C}$, 45 s
Hold at 4 $^{\circ}\text{C}$			

▲ CRITICAL STEP The initial 5 min incubation at 72 $^{\circ}\text{C}$ is critical for the success of the amplification reaction. This is because (i) transposed DNA contains nicks and overhangs that must be filled in prior to denaturation (Fig. 1b) and (ii) the polymerase enzyme in the NEBNext Ultra II Q5 2 \times Master Mix is a hot-start polymerase that becomes active at 45 $^{\circ}\text{C}$.

- 27 Remove tubes from the thermocycler, and store on ice. Proceed to the next step immediately.
- ▲ CRITICAL STEP** Additional cycles of amplification will be performed directly in this reaction tube, with the same reagents, so it is critical that samples remain cold and that the next step is performed immediately.

Library quantification and amplification ● Timing 3 h

- 28 Determine how the library concentrations will be quantified. Steps 29–32 detail an optimized method of quantification using the NEBNext Library Quant Kit, although other comparable products could be utilized following the manufacturer's recommendations. When preparing very few ATAC-seq libraries (fewer than four), an alternative quantification method using the Qubit fluorometer (Invitrogen) may prove faster, though less precise (Box 2).
- 29 Make a sufficient volume of 1 \times NEB Dilution Buffer by diluting the 10 \times NEB Dilution Buffer in water (~100 μL per sample).
- 30 Dilute 1 μL of pre-amplified sample with 99 μL of 1 \times NEB Dilution Buffer for a 1:100 dilution, and mix thoroughly.
- 31 Make 10 μL qPCR reactions using the NEBNext Library Quant Kit in a 384-well plate. Run all samples, the four standards and a no template control in technical duplicate.

Library quantification qPCR mix

Reagent	Volume per sample (μL)
NEBNext Library Quant Master Mix (with primer)	6
Diluted pre-amplified sample OR standard OR H ₂ O control	2
UltraPure distilled H ₂ O	2
Total volume	10

Box 2 | An alternative method for determining ATAC-seq library concentration after initial barcoding PCR

We recommend using qPCR to determine the concentration of ATAC-seq libraries. This provides an accurate estimate of how many additional cycles of amplification should be performed to obtain an optimally amplified library. This optimal amplification paradigm ensures that enough DNA is available for sequencing but avoids overamplification, which can introduce bias. It is also possible to estimate the number of additional cycles required via qPCR without a standard curve, as demonstrated previously¹⁶⁶.

However, as the additional number of cycles to amplify is ultimately an estimate, a Qubit fluorometer can be used in lieu of qPCR. We find that these estimates are best for samples using 25,000–100,000 nuclei or cells. For samples with <25,000 cells or nuclei, we strongly recommend using qPCR quantification, as the Qubit is often inaccurate.

To perform Qubit library quantification, after the library barcoding PCR in Step 27, use 1 µL of undiluted product as input to the Qubit dsDNA HS Assay Kit following manufacturer instructions. Determine the concentration, and estimate the number of additional cycles required using the equation below (see Supplementary Fig. 5 for data detailing how it was derived). We recommend performing two individual Qubit readings per ATAC-seq reaction for increased accuracy.

$$\text{Cycles} = -5.7 \times \log_{10} \left(\text{Qubit concentration} \left(\frac{\text{ng}}{\text{L}} \right) \right) + 6.7$$

Round the number of cycles to the nearest whole integer. On average, Qubit readings underestimate library concentration. This problem can be exacerbated at very low library concentrations. See an example calculation in Box 3.

- 32 Seal the plate, mix thoroughly by vortexing, pulse spin to collect sample at the bottom of each well and run according to the following cycling conditions:

Cycle no.	Denature	Anneal and extend
1	95 °C, 1 min	
2–36 (35 cycles)	95 °C, 15 s	63 °C, 45 s

- 33 After the qPCR is complete, use the standard curve to determine the concentration of DNA in the pre-amplified sample. New England Biolabs provides an online tool to assist in this calculation if needed at <http://nebiocalculator.neb.com/>. We have not found it necessary to adjust the resultant concentration on the basis of the average library fragment size. Typical library concentrations are 0.7–2 nM when using ~50,000 cells as input. However, the concentration will depend on many factors including input cell or nucleus quantity, viability and the proportion of cells or nuclei lost during processing.

? TROUBLESHOOTING

- 34 Use the concentration obtained in the previous step to estimate how many additional cycles are required to obtain a final yield of 240 fmol after column-based cleanup (20 µL at 12 nM). Round the fractional cycle number to the nearest whole integer. This PCR is highly efficient, so we assume a perfect doubling in concentration with each cycle. Note that we add one extra cycle in the below equation to ensure that sufficient library concentrations are obtained for all samples. See Box 3 for an example calculation.

$$\log_2 \left(\frac{240 \text{ fmol}}{49 \mu\text{L} \times [\text{PreAmp nM}]} \right) + 1 = \text{Cycles}$$

- 35 Place the tubes containing the pre-amplification reaction (now containing 49 µL) back in a thermocycler without addition of any more reagents, and run the required number of additional cycles for each individual sample.

Cycle no.	Denature	Anneal	Extend
1	98 °C, 30 s		
2 or more cycles (may vary between samples)	98 °C, 10 s	65 °C, 30 s	65 °C, 45 s
Final cycle		65 °C, 3 min	
Hold at 4 °C			

■ **PAUSE POINT** Amplified libraries can be stored at –20 °C overnight.

Box 3 | Example calculation for additional library amplification cycles

Here we detail an example calculation for how to determine the additional number of cycles either by qPCR quantification (recommended) or by Qubit reading. These two examples represent the results of qPCR and Qubit performed on the same pre-amplified ATAC-seq library. The 'Additional cycles' column gives the number of cycles that would be used in Step 35.

Measurement type	Concentration	Fractional cycles	Additional cycles
qPCR (nM)	0.34	4.85	5
Qubit (ng/ μ L)	1.5	5.69	6

- 36 Remove the PCR tubes from the thermocycler, and transfer the 49 μ L of amplified sample to a clean prelabeled 1.5 mL LoBind tube.
- 37 Add 245 μ L of DNA Binding Buffer (from the Zymo DNA Clean and Concentrator-5 kit) to each tube, and mix well by pipetting.
▲ CRITICAL STEP We strongly recommend the use of different kits/reagents for the cleanup of pre- and post-amplification products so as to not cross-contaminate post-amplification product into pre-amplification samples.
- 38 Transfer each sample, mixed with the DNA Binding Buffer, to a Zymo-Spin Column in a collection tube. Centrifuge at RT for 30 s at 10,000g, and discard the flowthrough.
- 39 Add 200 μ L of DNA Wash Buffer to the column, and centrifuge at RT for 30 s at 10,000g.
- 40 Repeat this wash for a total of two wash steps.
- 41 Perform a final 'dry spin' after the second wash step to remove any traces of residual wash buffer from the column membrane. To do this, remove any flowthrough from the collection tube and centrifuge the column and collection tube at RT for 1 min at >13,000g.
- 42 Transfer the column to a clean prelabeled 1.5 mL LoBind tube. Pipette 21 μ L of Elution Buffer directly onto the column membrane, and wait for 1 min.
- 43 Centrifuge the column at RT for 1 min at 13,000g to elute the DNA. This elution volume typically results in 20 μ L of product.
■ PAUSE POINT This solution can be stored at -20°C for as long as necessary.

Final library concentration determination ● Timing 2 h

▲ CRITICAL Here we describe the final library quantification by qPCR, which we find to be the most reliable method for determining the concentration of ATAC-seq libraries. Alternatively, Qubit can also be used to obtain an estimated library concentration for pooling samples.

- 44 Make a sufficient volume of 1 \times NEB Dilution Buffer by diluting the 10 \times NEB Dilution Buffer in water (~220 μ L per sample).
- 45 Dilute the samples from Step 43 4,000-fold so that they fall within the concentration range of the standards (0.01 pM to 10 pM). To do this, first dilute 40 \times by adding 0.5 μ L library to 19.5 μ L of 1 \times NEB Dilution Buffer. Mix well and then dilute 2 μ L of this 40 \times diluted mixture into 198 μ L of 1 \times NEB Dilution Buffer for a 100 \times dilution, creating a combined 4,000 \times dilution.

? TROUBLESHOOTING

- 46 Make 10 μ L qPCR reactions using the NEBNext Library Quant Kit in a 384-well plate. Run all diluted samples, the four standards and a no template control in technical duplicate.

Library quantification qPCR mix

Reagent	Volume per sample (μ L)
Primer + NEBNext Library Quant Master Mix	6
Diluted pre-amplified sample OR standard OR H ₂ O control	2
UltraPure distilled H ₂ O	2
Total volume	10

- 47 Seal the plate, mix thoroughly by vortexing, pulse spin to collect sample at the bottom of each well and run according to the following cycling conditions:

Cycle no.	Denature	Anneal and extend
1	95 °C, 1 min	
2–36 (35 cycles)	95 °C, 15 s	63 °C, 45 s

- 48 After the qPCR is complete, use the standard curve to determine the concentration of DNA in each of the final ATAC-seq libraries. New England Biolabs provides an online tool to assist in this calculation if needed at <http://nebiocalculator.neb.com/>. We have not found it necessary to adjust the resultant concentration based on the average library fragment size. Ideal library concentrations are 15–25 nM but can range from 2 to 100 nM, depending on the accuracy of the estimation of additional required PCR cycles.
- 49 If desired, dilute each library to 8 nM with UltraPure distilled water (or as low as 2 nM if needed). This facilitates pooling at equimolar quantities for sequencing. Samples are now ready for high-throughput sequencing.
- 50 We recommend using low-depth sequencing to check sample quality prior to high-depth sequencing. This will allow for calculation of the TSS enrichment score, observation of the fragment size distribution and determination of the percent of reads mapping to mitochondrial DNA. See the ‘Quality control of ATAC-seq libraries’ section of ‘Experimental design’ for more details. The PEPATAC pipeline can be used to calculate the TSS enrichment score. Alternatively, a Bioanalyzer trace can be used to roughly estimate library quality and concentration.

? TROUBLESHOOTING

- 51 After quality assessment, libraries should be sent for high-depth sequencing. If using a sequencing core facility, ask for the following cycling parameters:

Read	Read 1	i7 index	i5 index	Read 2
Purpose	ATAC-seq fragment	Sample barcode	Sample barcode	ATAC-seq fragment
Length	≥36 ^a	8	8	≥36 ^a

^aSee ‘Experimental design’ for ATAC-seq applications that may require longer reads.

Troubleshooting

Troubleshooting advice can be found in Table 5.

Table 5 | Troubleshooting table

Step	Problem	Possible reason	Solution
1	The cell or nuclei pellet is hard to visualize after centrifugation	Nuclei and certain cell types can be hard to properly pellet. Pellets of 50,000 cells, even for small cell types such as B cells, should be easily visualized	Add sterile BSA to a final concentration of 0.5% (wt/vol) or Tween-20 to a final concentration of 0.1% (wt/vol) to help cells or nuclei pellet properly
33	There is no amplification of the ATAC-seq libraries	Incorrect barcode/adaptor sequences were used when performing the barcoding step	Ensure that compatible barcodes were added, as detailed in Supplementary Table 2
	There is no amplification of the ATAC-seq libraries	Excess ethanol from the column may not have been properly removed using a dry spin prior to elution of transposed DNA	Ensure that the additional dry spin after the second wash step of the DNA Clean and Concentrator-5 kit is performed
45	The ATAC-seq libraries are outside the standard curve of the NEB Quant kit	The libraries may be either too dilute or too concentrated	We make dilutions in the 2,000–4,000-fold range. Adjust the dilutions of the final libraries as needed
50	Low-depth sequencing shows a low TSS enrichment score	Low signal-to-background ratios are often caused by unhealthy or otherwise non-ideal input material	Consider pretreating cells with DNase or using flow cytometry to sort viable cells
	There is no nucleosomal periodicity in the Bioanalyzer traces	Not all ATAC-seq libraries show nucleosomal periodicity on Bioanalyzer	This may not be a problem. See ‘Quality control of ATAC-seq libraries’

Timing ● Timing information is for ~12 samples.

Steps 1–19, transposition, 2.5 h
 Steps 20–27, barcoding of transposed fragments, 30 min
 Steps 28–43, library quantification and amplification, 3 h
 Steps 44–49, final library concentration determination, 2 h
 Steps 50–51, low- and/or high-depth sequencing of library, variable

Anticipated results

Brief overview of ATAC-seq analysis

ATAC-seq analysis can be broadly divided into three main stages: (1) data processing, (2) peak calling, merging and insertion counting, and (3) downstream analysis. These major steps are detailed in Fig. 4. *Data processing* refers to all the steps required to take the raw FASTQ files from the sequencer and prepare them for alignment to the genome and adjustment for the Tn5 offset (Supplementary Fig. 6; described below). Following this, the cleaned and adjusted reads are then used for *peak calling* on a sample-by-sample basis. These peaks are then merged across all samples that are being compared in the experiment, forming the union peak set, and the number of insertion events in each peak of the union peak set is counted across all samples. This insertion counts matrix and union peak set can be used for a variety of *downstream analyses*, including differential accessibility analysis to determine peaks that are significantly more or less accessible and motif enrichment on relevant sets of peaks. These stages, along with detailed descriptions of possible analysis tools, are described in greater detail below. In general, we recommend the use of the PEPATAC pipeline⁷³ as it performs many of the steps described below automatically—however, each use case may require tailoring of the precise tools used.

Computational resources needed to process ATAC-seq data

In general, the quantity of data generated from ATAC-seq libraries is too large to be analyzed on a standard personal laptop or desktop in a timely manner. We recommend the use of a high-performance computing environment, available at many institutions. The analysis pipelines depicted here can also be accomplished on smaller dedicated servers with fewer computational resources, for example, 32 cores and 128 GB of RAM. If a high-performance computing environment is not available, such services can also be purchased on a pay-per-use basis from cloud computing providers such as Amazon Web Services.

Pipelines to process ATAC-seq data

ATAC-seq data analysis consists of a variety of different steps, which are detailed in Fig. 4 and in the sections below. Often, each step requires a different tool and it can be complex for novice users to navigate between different input and output files. For these reasons, a variety of different pipelines have been developed for ATAC-seq data analysis, combining several or all of the steps of analysis together into an easy-to-use package. With these pipelines, the user specifies the input data and a predetermined set of analytical steps are performed to create the desired output files, including important QC metrics. Pipelines are highly recommended for beginners and experts alike, as they are robust, easily portable and easy to replicate. Several options exist, including, PEPATAC⁷³, the ENCODE ATAC-seq pipeline and nf-core^{71,84}. More ATAC-seq pipelines and analysis protocols are reviewed by Smith et al.⁹⁰. In general, most pipelines will take raw FASTQ files as input, calculate QC metrics, clean up and align the reads and provide ATAC-seq peaks as the final output. However, some pipelines do provide other features such as differential accessibility testing and motif enrichment determination. In Table 4, we compare the capabilities of three of the most commonly used pipelines. Each pipeline has its own advantages and disadvantages. For example, the ENCODE ATAC-seq pipeline was used to generate all of the ATAC-seq data files available from the ENCODE consortium, making it a good choice for users who aim to compare directly with those datasets. PEPATAC and nf-core provide a high degree of flexibility in choosing different parameters for analysis, without having to directly interface with each tool. PEPATAC was used for the analysis of the data presented in this protocol.

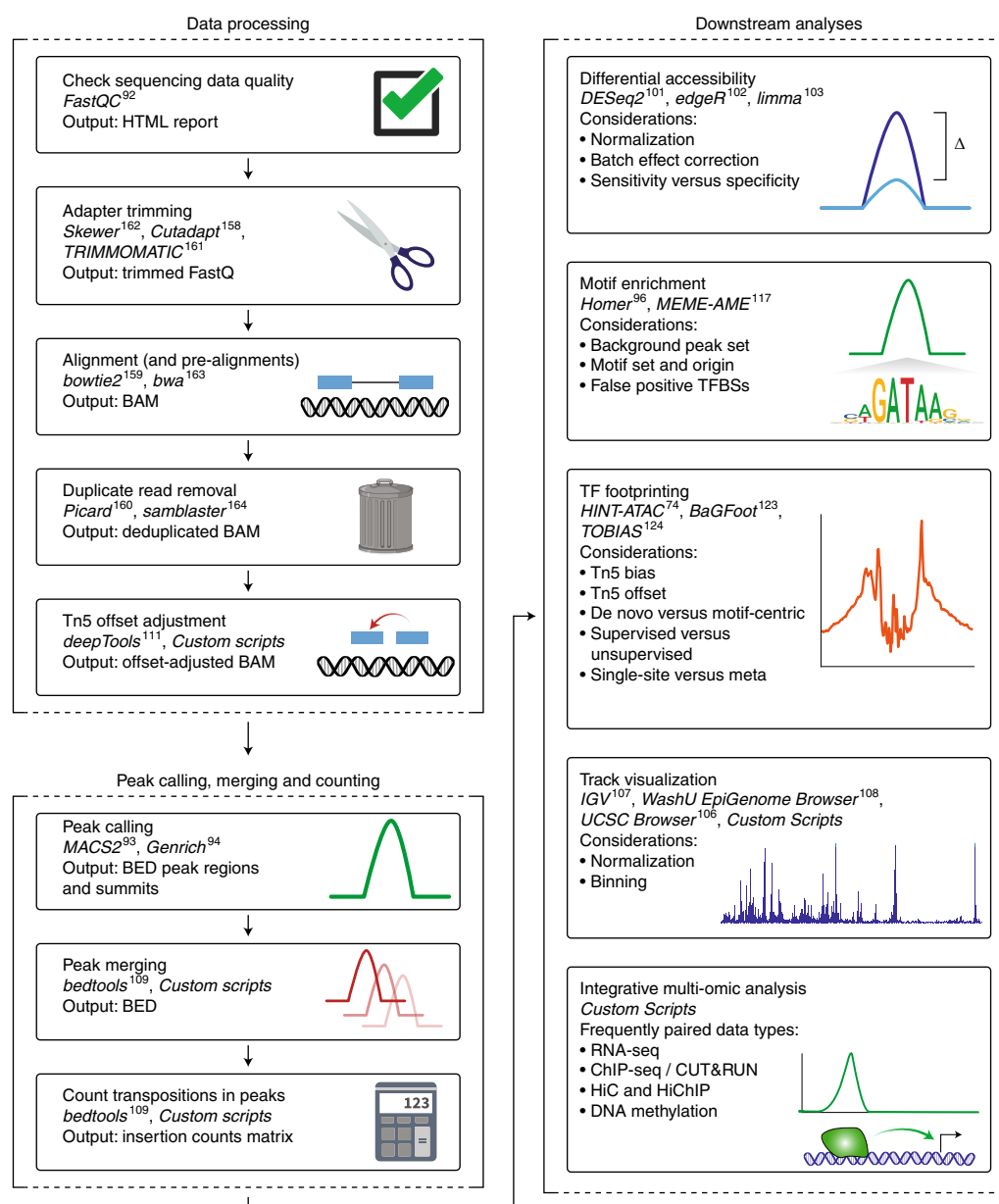


Fig. 4 | Overview of the steps of ATAC-seq data analysis. A schematic overview of the necessary steps in the ATAC-seq data analysis workflow, starting with the raw FASTQ files generated by the sequencer. Broadly, ATAC-seq analysis is composed of three major steps: (i) data processing that cleans and aligns the raw reads, (ii) peak calling, merging and insertion counting that determines the locations of Tn5 accessible chromatin and the relative signal within each accessible region, and (iii) the downstream analysis that can help assign putative functions and pathways to the called peaks. Each step, designated by a box, includes the possible bioinformatics tools that can be used to perform the analysis, as well as the expected output file type that should be utilized for the next step. Many pipelines can manage several aspects of this workflow automatically (Table 4).

Data processing after deep sequencing

The final output of the ATAC-seq protocol is typically FASTQ-formatted sequencing data files for each sample that was transposed, with each library having a read 1 and read 2 file, corresponding to the paired-end reads of either side of the ATAC-seq fragment (Fig. 1b). Each of these reads results from a unique Tn5 transposition event. The overall sequencing quality of each library can be checked using *FastQC*⁹¹, which provides an HTML report with per-base quality scores, as well as any observed biases in the data.

Standard ATAC-seq pipelines will take the FASTQ files as input and perform a series of QC and data cleaning steps followed by alignment to a reference genome. First, any bases corresponding to

the ATAC-seq adapter sequence will be removed (Fig. 4). This is important because ATAC-seq fragments can be shorter than the sequencing read length, resulting in read-through into the adapter sequence at the end of the read. Because this exogenous sequence interferes with proper alignment, it must be removed. Reads originating from mitochondrial DNA, which result from transposition into the unchromatinized mitochondrial genome and are considered ‘contaminating’, should be removed, either before full genome alignment or after. In general, with the Omni-ATAC protocol, we expect to see <15% of all reads mapping to the mitochondrial genome. Higher percentages may indicate over-lysis or insufficient removal of mitochondria during the wash step with ATAC-seq Wash Buffer (Step 7) and may require the libraries to be sequenced at higher depth to have sufficient unique reads mapping to the nuclear genome. The remaining reads are mapped to the reference genome and filtered for low-quality alignment.

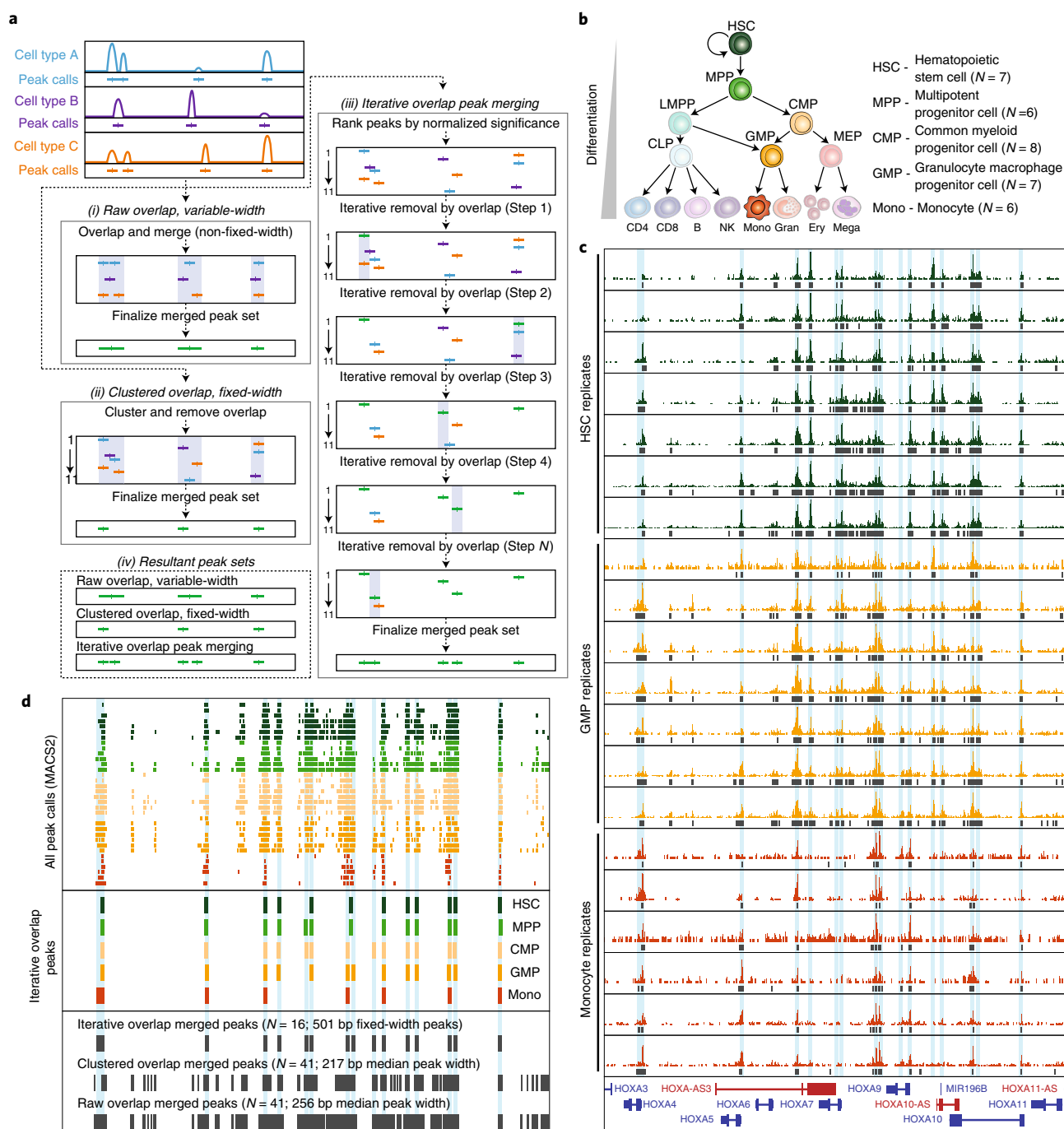
Next, PCR duplicates are marked and removed on the basis of identical start and end positions of the sequence fragment (Fig. 4). In general, we find that ATAC-seq libraries generated with this protocol from 50,000 cells require minimal amplification (three to seven extra cycles in Step 34). This level of amplification is associated with a library complexity that typically results in <10% of aligned reads marked as duplicates; however, it is important to note that the percent of reads marked as duplicates depends on many factors, including total sequencing depth, and we do not use this metric as a way to determine library quality. For example, samples with high levels of dead cells will result in libraries with high library complexity and low duplicate read percentages because their DNA is largely dechromatinized and thus is fully accessible to the transposase; however, this would not be considered a high-quality ATAC-seq library. Nevertheless, a high percentage of duplicate reads (>25%), especially when libraries are made from 50,000 cells and sequenced to ~10 million read pairs, may suggest that the transposition was inefficient and could reflect poor or insufficient input material. We find that most ATAC-seq libraries have >80% of the total reads mapping to the nuclear genome, after this series of QC steps. As described above, the TSS enrichment score is the single most important QC metric and should be recalculated on the deep sequencing data (Fig. 3a–c).

Finally, the ATAC-seq fragments are adjusted for the ‘Tn5 offset’. This Tn5 offset accounts for the molecular mechanism of binding and transposition by the Tn5 transposase dimer. Tn5 binds to DNA as a homodimer, with a 9 bp section of DNA between the two Tn5 molecules. Thus, each homodimer binding event creates two insertions, 9 bp apart, and the central 9 bp are duplicated and present on each of the corresponding fragments. The true center point of this Tn5 binding event is in the center of this 9 bp region. To account for this, the field has adopted the convention of adjusting the start position of each read by adding 4 bp to plus-stranded insertions and –5 bp to minus-stranded insertions³² (Supplementary Fig. 6).

Peak calling, merging and counting and annotation

After alignment, the next analysis step is to identify ‘peaks’ of chromatin accessibility, i.e., regions of the genome that are enriched for Tn5 insertions, suggesting that many cells in that sample had Tn5-accessible chromatin at that site. The location, and in some respects the magnitude, of those peaks can be informative in understanding the gene regulatory landscape.

Many programs exist to identify these peak regions, although MACS2⁹² and, more recently, Genrich⁹³ are two of the most frequently used (Fig. 4). HMMRATAC⁹⁴ was specifically designed for calling peaks in ATAC-seq data, and Genrich offers an ATAC-seq specific mode as well. As MACS2 is the most commonly used peak caller by ATAC-seq pipelines (Table 4), we focus on it here. Model-based analysis of ChIP-seq (MACS) is a popular peak calling tool for a variety of datasets, including ChIP-seq, ChIC-seq, DNase-seq and ATAC-seq. MACS2 calls peaks using a sliding window method, in which a user-provided window size is used to slide across the genome to find those regions with a fold enrichment (also user provided) relative to the background signal. Because MACS2 was not designed specifically for ATAC-seq, certain parameter adjustments are crucial. For example, to accurately represent the individual Tn5 insertion events in MACS2, the ‘shift’ and ‘extsize’ parameters should be used to ensure that the single-base position that represents the Tn5 insertion is directly in the middle of the ‘read’ provided to MACS2. To this end, the PEPATAC pipeline described above, performs peak calling with MACS2 with the following settings as default: `--shift -75 --extsize 150 --nomodel --call-summits --nolambda --keep-dup all -p 0.01`. It is important to note that peak calling in MACS2 is strongly affected by sequencing depth. Thus, the more reads in a dataset, the more peaks will be called and the significance of each peak will be greater. Therefore, it is not appropriate to compare raw peak calls across samples. In the past, we have advocated for the use of a normalized



peak score that can be obtained by normalizing the individual peak score ($-\log_{10}P$) to the sum total of all peak scores from the peaks identified in that sample³⁸.

Peak sets should first be identified on a sample-specific, and possibly technical-replicate-specific, basis and should then be merged between all samples that will be compared, creating what we refer to as the union peak set (Fig. 5). For example, if comparing cells treated with a compound to cells treated with a vehicle control, peaks should first be called for each technical replicate, then merged between all technical replicates in the group (e.g., all vehicle controls) and finally between the two experimental groups (e.g., treatment and vehicle). While this peak merging sounds straightforward, there are many ways to call and merge peaks, and the field has not yet come to a consensus on which is the most appropriate. We advocate for the use of non-overlapping fixed-width 501 bp peaks,

Fig. 5 | Schematic of peak merging strategies and the resulting merged peak sets. a, Schematic of the three possible peak merging options that have been frequently used for ATAC-seq data: raw overlap with variable-width peaks, clustered overlap with fixed-width peaks and iterative overlap with fixed-width peaks. This panel has been directly reproduced from ref. ¹⁶⁵. (i) In the raw overlap with variable-width peaks approach, any peaks that overlap with each other are merged together into a single, larger peak. This type of peak merging approach is implemented using the bedtools merge command, and will result in peaks that are of variable widths and often span multiple distinct regulatory elements. In the example shown in **d**, the raw overlap approach results in 41 peaks with a median peak width of 256 bp (± 408 bp standard deviation). (ii) In the clustered overlap with fixed-width peaks approach, clustered peaks are taken together and a single winner is chosen among them. This is typically implemented using the bedtools cluster command. The resulting merged peak set contains fixed-width peaks and has a tendency to underrepresent regulatory elements that are located in close proximity. In the example shown in **d**, the clustered overlap approach results in 41 peaks with a median peak width of 217 bp (± 326 bp standard deviation). (iii) In the iterative overlap with fixed-width peaks approach, first introduced in ref. ³⁸, fixed-width peaks are first ranked by their normalized significance. Once ranked, the most significant peak is retained, and any peaks directly overlapping with that peak are removed. This ranking and removal are iterated until there are no more overlapping peaks. The resulting merged peak set contains fixed-width peaks. In the example shown in **d**, the iterative overlap approach results in 16 peaks (after filtering by normalized significance) with a fixed peak width of 501 bp (0 bp standard deviation). (iv) Comparison of the resulting merged peak sets made using methods i–iii. **b**, Diagram of the hematopoietic differentiation hierarchy; to the right is number of samples used in **c** and **d** for each cell type. **c**, ATAC-seq signal tracks of the data from three distinct hematopoietic cell types from Corces et al. ³⁷. MPP and CMP data were excluded to improve figure legibility. Each track represents a different human donor. MACS2 peak calls are shown as black boxes below each signal track. **d**, Comparison of the MACS2 peak calls and the peak merging approaches for the tracks shown in **c**. Top: all MACS2 peak calls from **c** colored by their respective cell type. Middle: the cell-type-specific peak sets derived from the first round of the iterative overlap approach are shown after merging peaks from each of the biological replicates. Bottom: final merged peak set for all biological replicates across all cell types using the three methods described in **a**. The number of resulting peaks and their summary statistics are shown. CLP, common lymphoid progenitor; Ery, erythroblast; Gran, granulocyte; LMPP, lymphoid-primed multi-potent progenitor; Mega, megakaryocyte; MEP, megakaryocyte-erythroid progenitor cell; NK, natural killer cell.

centered on the peak summit, as a standard because most DNA regulatory elements are <500 bp in length³⁸. Fixed-width peaks make downstream computation easier because peak length does not need to be normalized, as compared with methods that use variable-width peaks. To merge peaks across samples, we advocate for the use of an iterative overlap procedure that maintains a fixed peak width while avoiding bias favoring deeper-sequenced libraries or differences in the number of replicates for a given sample type. In the iterative overlap, peaks are first ranked and filtered by their normalized significance (based on MACS2 scoring), then the most significant peak is retained, and any overlapping peaks are removed, iteratively, until no overlapping peaks remain. Iterative overlapping avoids two common problems of peak merging: (1) the creation of ever-larger peaks when multiple samples are compared with each other, since each sample-specific peak set will not overlap exactly, and (2) the loss of sensitivity that occurs when clustered overlap techniques are used, which take all the peaks that cluster together and keep a single winner to summarize the cluster. To facilitate the use of this iterative overlap peak merging approach with any set of MACS2 peak calls, we have made the peak merging script used in previous studies³⁸ available (https://github.com/corceslab/ATAC_IterativeOverlapPeakMerging).

Once a union peak set has been created, the number of transposition events per peak per sample is compiled into an insertion counts matrix. This raw matrix should be normalized for the total number of reads in peaks and then log transformed, though these operations are typically performed as part of standard differential analysis workflows. Peak functionality can partially be inferred by annotating peaks to the nearest gene. A variety of tools exist to perform this analysis, including HOMER (HOMER annotatePeaks.pl)⁹⁵, ChIPseeker⁹⁶ and ChIPpeakAnno⁹⁷. While the nearest gene provides a simple way to annotate peak function, it is important to remember that distal regulatory elements can act across large genomic distances and that the nearest gene often is not the true target for every peak^{38,98}.

Assessing the consistency of samples and technical replicates

After an insertion counts matrix has been obtained, the correlation between technical or biological replicates should be assessed. The precise Pearson correlation coefficient expected of biological replicates is largely dependent on the origin of the replicates. For example, one might expect CD4⁺ T cells isolated from three individual mice of the same inbred background to have a Pearson correlation value of >0.9, whereas the same cell type isolated from three different human donors might have a Pearson correlation value of >0.8 due to the larger natural variation between humans compared with inbred mice housed together. Regardless of the sample origin, we expect technical replicates (i.e., two ATAC-seq reactions performed side by side at the same time using different aliquots of the same starting material) to have a Pearson correlation >0.9 and ideally >0.95. These relationships between samples are typically visualized using a one-to-one plot or a Pearson correlogram.

Differentially accessible regions

Identification of differentially accessible regions (DARs) can be performed in a variety of different ways (Fig. 4) that have been well discussed previously⁹⁹. The choice of which workflow and tool to use for differential accessibility testing should be based on the goals of the project, and the amount of validation or follow-up analysis that will be performed. In general, the results from Gontarz et al.⁹⁹ suggest that DESeq and DESeq2¹⁰⁰ are the best choice for researchers who want to minimize false positives, i.e., minimize the false discovery rate. This may be appropriate for applications that will have little orthogonal validation and are designed to nominate possible high-likelihood candidate regions for further study. Conversely, for applications that aim to detect as many potential DARs as possible with tolerance to false positives, edgeR¹⁰¹ or limma¹⁰² are a good choice for high sensitivity, with limma giving the best results for regions with lower signal, such as distal enhancers. High-sensitivity techniques may be useful for studies leveraging integrative analysis with other sequencing techniques, which could lend additional confidence to DARs. An in-depth discussion of several different applications of ATAC-seq data, and the variety of tools that are designed to accomplish them, is also reviewed by Yan et al.¹⁰³.

Batch effects

As with most omics methods, batch effects can occur at multiple steps throughout the ATAC-seq library generation process and may impact downstream analyses, especially when comparing ATAC-seq datasets obtained at different timepoints or from different groups. In general, best practice rules should be used when generating ATAC-seq data within a specific experimental context to limit variation due to technical artifacts. However, some sources of batch effects may be challenging to control for as they can also originate from upstream sample attributes, for example, the post-mortem interval of human tissues. To remove these sources of unwanted variation between different ATAC experiments, one can use the built-in batch correction tools from edgeR or DESeq2⁹⁹. Both of these programs will allow for the inclusion of batch as a covariate in the experimental design. In edgeR, the batch is regressed out using a negative binomial generalized linear model for the batch and experimental conditions and performing the likelihood test as a generalization of the paired samples *t* test¹⁰¹. In a similar approach, DESeq2 also fits a negative binomial generalized linear model and uses the Wald test to determine the significance of the experimental condition¹⁰⁰. An alternate approach to these two programs is RUVseq¹⁰⁴, which is predominantly used for RNA-seq data, but has also been applied to ATAC-seq data⁹⁹. It performs factor analysis on the upper-quartile normalized counts using residuals calculated by edgeR¹⁰⁴.

Track visualization

Regardless of the workflow used to identify DARs, we recommend visual inspection of high-importance peaks using properly normalized sequencing tracks. This ensures that the difference identified through differential analysis is in line with expectation. A common way to do this is to create normalized bigWig (.bw) files that can be uploaded to genome browser tools such as the UCSC Genome Browser¹⁰⁵, Integrative Genomics Viewer¹⁰⁶ or the WashU Epigenome Browser¹⁰⁷. BigWig files provide a streamlined way to visualize genome-wide alignments and, in the case of ATAC-seq data, genome-wide patterns in chromatin accessibility. Normalization can be performed using bedtools genomecov^{108,109} or deepTools bamCoverage¹¹⁰ by applying a scale factor based on the number of reads in TSS regions. This approach is analogous to normalizing based on reads in peaks, enabling simultaneous normalization for both sequencing depth and data quality. We recommend using reads in TSS regions because these regions are invariant across samples, whereas use of reads in peaks requires renormalization any time the peak set changes. We also recommend using a standardized bin size when creating the bigWig files. The larger the bin size, the lower the resolution and the smaller the resulting bigWig file. We recommend a bin size of 100 bp for most applications.

Motif enrichment analysis

Motif enrichment analysis is based on searching for TF binding motifs within a given peak set. Motif information, from both experimental observations as well as computational predictions, has been collated in a variety of large databases, including JASPAR¹¹¹, CIS-BP¹¹² and ENCODE¹¹³. Each motif is stored as a position-weight matrix, and a variety of tools are able to scan the DNA sequence of each peak for the presence of each motif, including HOMER⁹⁵, TFBSTools¹¹⁴, motifmatchr¹¹⁵ and MEME¹¹⁶. Enrichment for the presence of a given motif within the peak set of interest compared with

a set of background regions implies that the given TF may be important in driving the accessibility at those peak regions. For example, when comparing differentiated with undifferentiated cells, these motif enrichment tools would identify motifs that are statistically enriched in the peaks gained after differentiation, compared with those shared between the differentiated and undifferentiated cells. This could provide a hypothesis about the TF(s) responsible for driving differentiation. Various statistical tests have been used to identify motif enrichment, including the hypergeometric test (HOMER) and the rank sum test (MEME). When performing these analyses, it is important to select a meaningful set of regions or peaks to serve as the background for enrichment testing. It is often not correct to use the entire genome as the background and a matched subset of peaks or the whole peak set should be used instead, depending on the particular application.

While powerful tools, it is important to remember that peak-scanning-based motif enrichment methods can contain many false positives, as multiple different motifs may be present within a peak, and often TFs from the same family share similar motifs. Additionally, motif enrichment is not direct evidence for TF binding. One way to increase the biological meaning of motif enrichment results is to refine potential TF mediators on the basis of their expression in the cell type or tissue of interest. Another way to analyze differences in motifs is to calculate the difference in observed versus expected accessibility at peak sets. This type of analysis is predicated on the idea that greater accessibility, i.e., insertion counts, is correlated with greater TF activity. Thus, a difference in fragment counts can be taken as a change in the activity of that TF. Two programs have been implemented to do this type of analysis: chromVAR¹¹⁷ and diffTF¹¹⁸. chromVAR, which was originally developed for single-cell ATAC-seq data, calculates an accessibility deviation across multiple conditions for each motif. diffTF, which is conceptually similar to chromVAR, was designed to work with bulk ATAC-seq data and can also take matched RNA-seq data as input, for its ‘classification’ mode to determine the mode of action for a given TF (i.e., repressor or activator)¹¹⁸. diffTF is run from within a Snakemake pipeline workflow and uses aligned BAM files as input.

TF footprinting analysis

TF regulation of the accessible peaks can also be determined by footprinting analysis. Footprinting analysis originates from the classic DNase footprint^{27,28}, and was later adapted to sequencing-based analysis (sometimes termed digital genomic footprinting¹¹⁹). Within the context of ATAC-seq, TF footprints would occur when a TF binds to DNA and prevents Tn5 insertions specifically at the DNA bases that are directly bound. This leaves a section of DNA, or a ‘footprint’, with a dearth of insertions compared with the adjacent nucleosome-free (and TF-free) regions (Fig. 4). Thus, the goal of footprinting algorithms is, first, to identify these locations of ‘dipped’ Tn5 insertions and, second, to determine which TF might have been bound there by examining the sequence of the central bases inferred to be directly bound by the TF. Identifying TF footprints can aid in the reconstruction of gene regulatory networks, and may be more specific than motif scanning in detecting the presence of a TF. However, there are many challenges to applying footprinting analysis to ATAC-seq data, and the interpretation can often be confounded by Tn5’s sequence bias²⁸. Classically, TF footprinting was designed to predict TF binding at a single site in the genome. This requires many fragments mapping to that particular locus to provide sufficient observations to identify the protected bases. In the context of ATAC-seq performed on 50,000 cells, it is often the case that insufficient depth is obtained to perform single-site footprinting, even at high sequencing depth. Because of this, many researchers have performed ‘meta’ footprinting where hundreds or thousands of genomic locations harboring the motif for a given TF are collated into a single footprint. It is important to note that this is not the classical use of footprinting and its results have more in common with motif enrichment than they do with single-site footprinting. As with other ATAC-seq methods, TF footprinting analysis should adjust fragments for the Tn5 offset^{32,120} (Supplementary Fig. 6) and account for the Tn5 insertion bias¹²¹. Several approaches for single-site footprinting, including HINT-ATAC⁷⁴, BaGFoot¹²² and TOBIAS¹²³, are discussed in Supplementary Note 4.

Nucleosome positioning

ATAC-seq data have also been used to look at nucleosome positioning⁶¹, although these techniques are still in development, and require very high sequencing depth to be accurately calculated. Nucleosome positioning refers to techniques that attempt to measure differences in the organization of nucleosomes at certain locations, including their position and occupancy¹²⁴. Changes in nucleosome profiles have been associated with changes in gene regulation and expression levels at a given

promoter¹²⁵. Classically, nucleosome positioning has been performed with MNase-seq data, and several tools have been developed for that type of data. However, NucleoATAC⁶¹ and HMMRATAC⁹⁴ were developed specifically for predicting nucleosome positions from ATAC-seq data. In general, for the sequencing depth of the average ATAC-seq library, applying nucleosome profiling techniques is not advised.

Integrating ATAC-seq with other omic profiling techniques

As mentioned in the introduction, ATAC-seq gives a general landscape of chromatin accessibility across the genome. As such, it can be interesting to compare the chromatin accessibility landscapes from ATAC-seq with other matched genomic, transcriptional or epigenomic profiles. Intersecting these datasets may help to interpret the function of a given ATAC-seq peak. For example, ATAC-seq can be intersected with H3K27ac ChIP-seq (or other ChIC techniques) with the expectation that a subset of ATAC-seq peaks would overlap with the active enhancers and promoters marked by H3K27ac. With a large enough dataset, cross-sample correlations between ATAC-seq chromatin accessibility and nearby gene expression can be used to identify peak-to-gene links that represent putative gene regulatory interactions. Such interactions could be orthogonally supported using chromosome conformation capture techniques such as HiC¹²⁶ or HiChIP¹²⁷.

Similarly, ATAC-seq provides a unique window into the effects of sequence changes in the noncoding genome. By combining ATAC-seq data with whole-genome sequencing data, one can identify regions of allelic imbalance whereby the two different alleles show differential accessibility, implying differential TF binding¹²⁸. This has been used to identify chromatin accessibility quantitative trait loci¹²⁸ and to begin assigning function to noncoding polymorphisms identified through GWASs^{53,129,130}. These types of analyses can help to assign relevance or function to specific putative regulatory elements identified through ATAC-seq.

Preparing for publication

ATAC-seq data included in publications should be deposited in publicly available repositories, such as the Gene Expression Omnibus, the Sequencing Read Archive or the European Nucleotide Archive. Instructions specific to each repository can be found on their websites. However, it is most common to include the following files: (1) all raw FASTQ files (read 1 and read 2) prior to QC and adapter trimming, (2) an insertion counts matrix for the union peak set of all samples analyzed in the study, including the genomic coordinates of the peaks, (3) normalized bigWig files that can be easily visualized on a genome browser and (4) a list of DARs if applicable. We also encourage researchers to include a supplementary table reporting the QC metrics, as detailed above, for each of their sequenced samples, especially the TSS enrichment score.

Data availability

The ATAC-seq datasets generated for the protocol optimizations detailed in Supplementary Figs. 1, 4 and 5 are available on the Gene Expression Omnibus under accession number [GSE188797](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE188797). The data used in Fig. 5 are taken from ref. ³⁸. All analyses were performed using the hg38 human genome.

Code availability

The source code for the iterative overlap is freely available at https://github.com/corceslab/ATAC_IterativeOverlapPeakMerging⁸⁸. All other ATAC-seq data analysis for the figures used in this protocol were generated using PEPATAC⁷³ with Bulker (container version 1.0.8), available at <http://pepatac.databio.org/en/latest/>.

References

1. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).
2. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
3. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
4. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein–DNA interactions. *Science* **316**, 1497–1502 (2007).
5. Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nat. Rev. Genet.* **13**, 840–852 (2012).

6. Nakato, R. & Sakata, T. Methods for ChIP-seq analysis: a practical workflow and advanced applications. *Methods* **187**, 44–53 (2021).
7. Schmid, M., Durussel, T. & Laemmli, U. K. ChIC and ChEC: genomic mapping of chromatin proteins. *Mol. Cell* **16**, 147–157 (2004).
8. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
9. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
10. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).
11. Wang, Q. et al. CoBATCH for high-throughput single-cell epigenomic profiling. *Mol. Cell* **76**, 206–216.e7 (2019).
12. Handa, T. et al. Chromatin integration labeling for mapping DNA-binding proteins and modifications with low input. *Nat. Protoc.* **15**, 3334–3360 (2020).
13. Harada, A. et al. A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.* **21**, 287–296 (2019).
14. Ku, W. L. et al. Single-cell chromatin immunocleavage sequencing (scChIC-Seq) to profile histone modification. *Nat. Methods* **16**, 323–325 (2019).
15. Zheng, X.-Y. & Gehring, M. Low-input chromatin profiling in *Arabidopsis* endosperm using CUT&RUN. *Plant Reprod.* **32**, 63–75 (2019).
16. Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. & Fazio, T. G. Profiling of pluripotency factors in single cells and early embryos. *Cell* **177**, 1319–1329 (2019).
17. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
18. Crawford, G. E. et al. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* **16**, 123–131 (2006).
19. Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5384 (2010).
20. Boyle, A. P. et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
21. Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877–885 (2007).
22. Cui, K. & Zhao, K. in *Chromatin Remodeling: Methods and Protocols* (ed. Morse, R. H.) 413–419 (Humana, 2012).
23. Schones, D. E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
24. Kelly, T. K. et al. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).
25. Minnoye, L. et al. Chromatin accessibility profiling methods. *Nat. Rev. Methods Primer* **1**, 1–24 (2021).
26. Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).
27. Galas, D. J. & Schmitz, A. DNase footprinting: a simple method for the detection of protein–DNA binding specificity. *Nucleic Acids Res.* **5**, 3157–3170 (1978).
28. He, H. H. et al. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11**, 73–78 (2014).
29. Sung, M.-H., Baek, S. & Hager, G. L. Genome-wide footprinting: ready for prime time? *Nat. Methods* **13**, 222–228 (2016).
30. Mieczkowski, J. et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat. Commun.* **7**, 11485 (2016).
31. Chereji, R. V., Bryson, T. D. & Henikoff, S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol.* **20**, 198 (2019).
32. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
33. Goryshin, I. Y. & Reznikoff, W. S. Tn5 in vitro transposition. *J. Biol. Chem.* **273**, 7367–7374 (1998).
34. Adey, A. et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119 (2010).
35. Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S. J. & Craig, N. L. DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. *Proc. Natl Acad. Sci. USA* **107**, 21966–21972 (2010).
36. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
37. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).

38. Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
39. Calderon, D. et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494–1505 (2019).
40. Naik, S. et al. Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. *Nature* **550**, 475–480 (2017).
41. Marco, A. et al. Mapping the epigenomic and transcriptomic interplay during memory formation and recall in the hippocampal engram ensemble. *Nat. Neurosci.* **23**, 1606–1617 (2020).
42. Li, D. et al. Chromatin accessibility dynamics during iPSC reprogramming. *Cell Stem Cell* **21**, 819–833 (2017).
43. Guo, J. et al. Chromatin and single-cell RNA-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell* **21**, 533–546 (2017).
44. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
45. Daugherty, A. C. et al. Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. *Genome Res.* **27**, 2096–2107 (2017).
46. Gury-BenAri, M. et al. The spectrum and regulatory landscape of intestinal innate lymphoid cells are shaped by the microbiome. *Cell* **166**, 1231–1246 (2016).
47. Liu, Q. et al. Chromatin accessibility landscapes of skin cells in systemic sclerosis nominate dendritic cells in disease pathogenesis. *Nat. Commun.* **11**, 5843 (2020).
48. Liu, Y. et al. Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. *Sci. Rep.* **8**, 15499 (2018).
49. Greenwald, W. W. et al. Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nat. Commun.* **10**, 2078 (2019).
50. Lee, J. et al. Activation of PDGF pathway links LMNA mutation to dilated cardiomyopathy. *Nature* **572**, 335–340 (2019).
51. Schmidl, C. et al. Combined chemosensitivity and chromatin profiling prioritizes drug combinations in CLL. *Nat. Chem. Biol.* **15**, 232–240 (2019).
52. Scharer, C. D. et al. Epigenetic programming underpins B cell dysfunction in human SLE. *Nat. Immunol.* **20**, 1071–1082 (2019).
53. Corces, M. R. et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).
54. Banovich, N. E. et al. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res.* **28**, 122–131 (2018).
55. Forrest, M. P. et al. Open chromatin profiling in hiPSC-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. *Cell Stem Cell* **21**, 305–318 (2017).
56. Chiou, J. et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398–402 (2021).
57. Nott, A. et al. Brain cell type-specific enhancer–promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
58. Liu, Q. et al. Genome-wide temporal profiling of transcriptome and open-chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. *Circ. Res.* **121**, 376–391 (2017).
59. Wapinski, O. L. et al. Rapid chromatin switch in the direct reprogramming of fibroblasts to neurons. *Cell Rep.* **20**, 3236–3247 (2017).
60. Denny, S. K. et al. Nf1b promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342 (2016).
61. Schep, A. N. et al. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* **25**, 1757–1770 (2015).
62. Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.* **15**, 3264–3283 (2020).
63. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).
64. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* **39**, 825–835 (2021).
65. Gao, W., Lai, B., Ni, B. & Zhao, K. Genome-wide profiling of nucleosome position and chromatin accessibility in single cells using scMNase-seq. *Nat. Protoc.* **15**, 68–85 (2020).
66. Lai, B. et al. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* **562**, 281–285 (2018).
67. Jin, W. et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528**, 142–146 (2015).
68. Takaku, M. et al. GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol.* **17**, 36 (2016).
69. Fujiwara, S., Baek, S., Varticovski, L., Kim, S. & Hager, G. L. High quality ATAC-seq data recovered from cryopreserved breast cell lines and tissue. *Sci. Rep.* **9**, 516 (2019).
70. Mulqueen, R. M. et al. Improved single-cell ATAC-seq reveals chromatin dynamics of in vitro corticogenesis. Preprint at *bioRxiv* <https://doi.org/10.1101/637256> (2019).

71. *nf-core/atacseq*. (nf-core, 2021). <https://doi.org/10.5281/zenodo.2634132>
72. ATAC-seq Data Standards and Processing Pipeline. *ENCODE*. <https://www.encodeproject.org/atac-seq/>
73. Smith, J. P. et al. PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments. *NAR Genomics Bioinform.* **3**, lqab101 (2021).
74. Bajic, M., Maher, K. A. & Deal, R. B. Identification of open chromatin regions in plant genomes using ATAC-seq. *Methods Mol. Biol.* **1675**, 183–201 (2018).
75. Deal, R. B. & Henikoff, S. The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat. Protoc.* **6**, 56–68 (2011).
76. Haines, J. E. & Eisen, M. B. Patterns of chromatin accessibility along the anterior-posterior axis in the early *Drosophila* embryo. *PLOS Genet.* **14**, e1007367 (2018).
77. Johnson, S., Nguyen, V. & Coder, D. Assessment of cell viability. *Curr. Protoc. Cytom.* **64**, 9.2.1–9.2.26 (2013).
78. Chen, X. et al. ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
79. Swanson, E. et al. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *eLife* **10**, e63632 (2021).
80. Corces, R. Isolation of nuclei from frozen tissue for ATAC-seq and other epigenomic assays. <https://doi.org/10.17504/protocols.io.6t8herw> (2019).
81. Polavarapu, V. K. et al. Profiling chromatin accessibility in formalin-fixed paraffin-embedded samples. *Genome Res.* <https://doi.org/10.1101/gr.275269.121> (2021).
82. Chin, H. G. et al. Universal NicE-seq for high-resolution accessible chromatin profiling for formaldehyde-fixed and FFPE tissues. *Clin. Epigenetics* **12**, 143 (2020).
83. Orchard, P., Kyono, Y., Hensley, J., Kitzman, J. O. & Parker, S. C. J. Quantification, dynamic visualization, and validation of bias in ATAC-seq data with ataqv. *Cell Syst.* **10**, 298–306 (2020).
84. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
85. Koohy, H., Down, T. A. & Hubbard, T. J. Chromatin accessibility data sets show bias due to sequence specificity of the DNase I enzyme. *PLoS ONE* **8**, e69853 (2013).
86. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
87. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
88. Granja, J. M. & Corces, M. R. ATAC_IterativeOverlapPeakMerging <https://doi.org/10.5281/zenodo.5903680> (2022).
89. Picelli, S. et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
90. Smith, J. P. & Sheffield, N. C. Analytical approaches for ATAC-seq data analysis. *Curr. Protoc. Hum. Genet.* **106**, e101 (2020).
91. A Quality Control tool for High Throughput Sequence Data. *Babraham Bioinformatics*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
92. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
93. Gaspar, J. M. *Genrich: Detecting Sites of Genomic Enrichment* (2021).
94. Tarbell, E. D. & Liu, T. HMMRATAC: a Hidden Markov ModelER for ATAC-seq. *Nucleic Acids Res.* **47**, e91 (2019).
95. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
96. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
97. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
98. Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).
99. Gontarz, P. et al. Comparison of differential accessibility analysis strategies for ATAC-seq data. *Sci. Rep.* **10**, 10150 (2020).
100. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
101. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
102. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
103. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker’s guide to ATAC-seq data analysis. *Genome Biol.* **21**, 22 (2020).
104. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
105. UCSC Genome Browser Home. <https://genome.ucsc.edu/index.html>
106. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

107. WashU Epigenome Browser. <http://epigenomegateway.wustl.edu/browser/>
108. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
109. genomcov. *bedtools* 2.30.0 documentation <https://bedtools.readthedocs.io/en/latest/content/tools/genomcov.html>
110. bamCoverage. *deepTools* 3.5.0 documentation <https://deeptools.readthedocs.io/en/develop/content/tools/bamCoverage.html>
111. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
112. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
113. Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* **42**, 2976–2987 (2014).
114. Tan, G. & Lenhard, B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556 (2016).
115. Fast Motif Matching in R. *motifmatchr*. <https://greenleaf.github.io/motifmatchr/index.html>
116. Bailey, T. L. et al. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
117. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
118. Berest, I. et al. Quantification of differential transcription factor activity and multiomics-based classification into activators and repressors: diffTF. *Cell Rep.* **29**, 3147–3159 (2019).
119. Hesselberth, J. R. et al. Global mapping of protein–DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
120. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
121. Martins, A. L., Walavalkar, N. M., Anderson, W. D., Zang, C. & Guertin, M. J. Universal correction of enzymatic sequence bias reveals molecular signatures of protein/DNA interactions. *Nucleic Acids Res.* **46**, e9 (2018).
122. Baek, S., Goldstein, I. & Hager, G. L. Bivariate genomic footprinting detects changes in transcription factor activity. *Cell Rep.* **19**, 1710–1722 (2017).
123. Bentsen, M. et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).
124. Singh, A. K. & Mueller-Planitz, F. Nucleosome positioning and spacing: from mechanism to function. *J. Mol. Biol.* **433**, 166847 (2021).
125. Lai, W. K. M. & Pugh, B. F. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nat. Rev. Mol. Cell Biol.* **18**, 548–562 (2017).
126. Belton, J.-M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
127. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
128. Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.* **48**, 206–213 (2016).
129. Gate, R. E. et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* **50**, 1140–1150 (2018).
130. Örd, T. et al. Single-cell epigenomics and functional fine-mapping of atherosclerosis GWAS loci. *Circ. Res.* **129**, 240–258 (2021).
131. Kia, A. et al. Improved genome sequencing using an engineered transposase. *BMC Biotechnol.* **17**, 6 (2017).
132. Green, B., Bouchier, C., Fairhead, C., Craig, N. L. & Cormack, B. P. Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mob. DNA* **3**, 3 (2012).
133. Ason, B. & Reznikoff, W. S. DNA sequence bias during Tn5 transposition. *J. Mol. Biol.* **335**, 1213–1225 (2004).
134. Lazarovici, A. et al. Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl Acad. Sci. USA* **110**, 6376–6381 (2013).
135. Dingwall, C., Lomonosoff, G. P. & Laskey, R. A. High sequence specificity of micrococcal nuclease. *Nucleic Acids Res.* **9**, 2659–2673 (1981).
136. Hörz, W. & Altenburger, W. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Res.* **9**, 2643–2658 (1981).
137. Wu, S. J. et al. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat. Biotechnol.* **39**, 819–824 (2021).
138. Patty, B. J. & Hainer, S. J. Transcription factor chromatin profiling genome-wide using uliCUT&RUN in single cells and individual blastocysts. *Nat. Protoc.* **16**, 2633–2666 (2021).
139. Carter, B. et al. Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). *Nat. Commun.* **10**, 3747 (2019).
140. Chang, P., Gohain, M., Yen, M.-R. & Chen, P.-Y. Computational methods for assessing chromatin hierarchy. *Comput. Struct. Biotechnol. J.* **16**, 43–53 (2018).

141. Mo, A. et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015).
142. Yang, A. C. et al. Dysregulation of brain and choroid plexus cell types in severe COVID-19. *Nature* **595**, 565–571 (2021).
143. Drokhyansky, E. et al. The human and mouse enteric nervous system at single-cell resolution. *Cell* **182**, 1606–1622 (2020).
144. Deal, R. B. & Henikoff, S. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev. Cell* **18**, 1030–1040 (2010).
145. Bhattacharyya, S., Sathe, A. A., Bhakta, M., Xing, C. & Munshi, N. V. PAN-INTACT enables direct isolation of lineage-specific nuclei from fibrous tissues. *PLoS ONE* **14**, e0214677 (2019).
146. Nuclei Isolation for Single Cell ATAC Sequencing. <https://support.10xgenomics.com/single-cell-atac/sample-prep/doc/demonstrated-protocol-nuclei-isolation-for-single-cell-atac-sequencing>
147. Lawler, A. J. et al. Cell type-specific oxidative stress genomic signatures in the globus pallidus of dopamine-depleted mice. *J. Neurosci.* **40**, 9772–9783 (2020).
148. Kiseleva, E. et al. A protocol for isolation and visualization of yeast nuclei by scanning electron microscopy (SEM). *Nat. Protoc.* **2**, 1943–1953 (2007).
149. Niepel, M., Farr, J. C., Rout, M. P. & Strambio-De-Castillia, C. Rapid isolation of functionally intact nuclei from the yeast *Saccharomyces*. Preprint at *bioRxiv* <https://doi.org/10.1101/162388> (2017).
150. Nott, A., Schlachetzki, J. C. M., Fixsen, B. R. & Glass, C. K. Nuclei isolation of multiple brain cell types for omics interrogation. *Nat. Protoc.* **16**, 1629–1646 (2021).
151. Fullard, J. F. et al. An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28**, 1243–1252 (2018).
152. Hauberg, M. E. et al. Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nat. Commun.* **11**, 5581 (2020).
153. Haines, J. ATAC-seq on nuclei from frozen, sliced, *Drosophila melanogaster* embryo halves. <https://doi.org/10.17504/protocols.io.kj5cuq6> (2017).
154. Steiner, F. A., Talbert, P. B., Kasinathan, S., Deal, R. B. & Henikoff, S. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res.* **22**, 766–777 (2012).
155. Han, M., Wei, G., McManus, C. E., Hillier, L. W. & Reinke, V. Isolated *C. elegans* germ nuclei exhibit distinct genomic profiles of histone modification and gene expression. *BMC Genomics* **20**, 500 (2019).
156. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
157. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
158. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).
159. Krueger, F. *Trim Galore*. (2021).
160. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
161. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
162. *broadinstitute/picard* (Broad Institute, 2021) <https://github.com/broadinstitute/picard>
163. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
164. Boyle, A. P., Guinney, J., Crawford, G. E. & Furey, T. S. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics* **24**, 2537–2538 (2008).
165. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
166. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015).

Acknowledgements

This work was supported by NIH R00-AG059918, U01-AG072573, P01-AG073082, UM1-HG012076 and a gift from the Ray and Dagmar Dolby Family Fund (to the Gladstone Institutes). F.C.G. is an Alan Kaganov Scholar. M.R.C. is additionally supported by the Farmer Family Foundation Parkinson’s Research Initiative and an American Society of Hematology Scholar Award.

Author contributions

All authors contributed to developing this protocol. All experiments were performed by H.M., F.C.G. and L.K. with supervision from M.R.C. The manuscript was written by F.C.G., L.K. and H.M. with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41596-022-00692-9>.

Correspondence and requests for materials should be addressed to M Ryan Corces.

Peer review information *Nature Protocols* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

Received: 1 June 2021; Accepted: 22 February 2022;

Published online: 27 April 2022

Related links**Key references using this protocol**

Corces, M. R. et al. *Nat. Methods* **14**, 959–962 (2017): <https://doi.org/10.1038/nmeth.4396>

Corces, M. R. et al. *Science* **362**, 6413 (2018): <https://doi.org/10.1126/science.aav1898>

Corces, M. R. et al. *Nat. Genet.* **52**, 1158–1168 (2020): <https://doi.org/10.1038/s41588-020-00721-x>