Chromatin accessibility profiling methods

Liesbeth Minnoye[1,2], Georgi K. Marinov[3], Thomas Krausgruber[4], Lixia Pan[5], Alexandre P. Marand[6], Stefano Secchia[7], William J. Greenleaf [3], Eileen E. M. Furlong[7], Keji Zhao[5], Robert J. Schmitz[6], Christoph Bock[4,8], and Stein Aerts[1,2]

Affiliations

1. VIB-KU Leuven Center for Brain & Disease Research, Leuven, Belgium.
2. KU Leuven, Department of Human Genetics KU Leuven, Leuven, Belgium.
3. Department of Genetics, Stanford University, Stanford, CA, USA.
4. CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria.
5. Laboratory of Epigenome Biology, Systems Biology Center, Division of Intramural Research, National Heart, Lung and Blood Institute, NIH, Maryland, USA.
6. Department of Genetics, University of Georgia, Athens, GA, U.S.A.
7. European Molecular Biology Laboratory (EMBL), Genome Biology Department, Meyerhofstrasse 1, 69117 Heidelberg, Germany.
8. Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria.

Abstract

Chromatin accessibility, or the compaction of the complex of DNA and associated proteins, is a widely studied characteristic of the eukaryotic genome. Since the 1970s, research on chromatin accessibility has been instrumental for studying genome regulation. As regulatory DNA is generally found accessible when it is operational, genome-wide profiling of chromatin accessibility can be used as experimental tool to identify all candidate genomic regions that underlie the regulatory state of a tissue or cell type. Multiple biochemical techniques have been developed to profile chromatin accessibility, which have yielded an extensive source of chromatin accessibility maps across a broad range of species, tissues, cell types and diseases. With the help of *cis*-regulatory sequence analysis and the adoption of single-cell chromatin accessibility profiling, insight into the key regulators underlying developmental and disease processes is rapidly increasing. Both bulk and single-cell methods are based on high-throughput sequencing, making computational analysis and bioinformatics tools invaluable for the exploration and interpretation of the generated data. We foresee exciting technological improvements including single-molecule, multi-omics, and spatial methods to bring further insight into the outstanding secrets of genome control.

1. Introduction

Chromatin accessibility refers to the level of physical compaction of chromatin, a complex formed by DNA and associated proteins consisting mainly of histones and DNA-binding transcription factors (TF)[1–3]. Although eukaryotic genomes are generally packed into nucleosomes, which comprise ~147 bp of DNA wrapped around an octamer of histones[4,5],

nucleosomal occupancy is not uniform across the genome, and varies across tissues and cell types. Nucleosomes are typically depleted at genomic locations that interact with transcriptional regulators (e.g. TFs), such as at enhancers, promoters and other regulatory elements; which thus present themselves as 'accessible' or 'open' chromatin[6–10]. Therefore, profiling chromatin accessibility on a genome-wide scale serves as an excellent tool to map putative regulatory elements in a cell type or cell state. Note that not only nucleosome positioning, but also chemical modifications of the chromatin, including DNA methylation (in mammals) and histone tail methylation and acetylation, are dynamic and change between different cell states. These modifications, which are often correlated with chromatin accessibility, can reflect specific functionalities of genomic regions in relation to the regulation of gene expression[11,12]. Initial changes in accessibility are due to the binding of TFs, which outcompete histones and recruit co-factors, including ATP-dependent chromatin remodelers[13,14]; or TFs that preferentially bind to their recognition sequence in nucleosomal DNA[15,16]. The binding of such "pioneer factors" can facilitate other TFs to co-bind and further stabilize the nucleosome depleted region and cooperatively regulate gene expression of target genes[17–19]. Consequently, the analysis of TF binding sites within accessible regions can bring insights into cell type specific master regulators and gene regulatory networks.

Changes in the chromatin landscape, as well as mutations in chromatin remodelers and in regulatory regions, have been linked to a range of traits and diseases[20–23]. In fact, many causal genome-wide association study (GWAS) variants are located in accessible regulatory elements[24] and TF-bound DNA harbors increased mutation rates since TFs and DNA repair enzymes compete for damaged regulatory regions[25,26]. In order to improve our understanding of chromatin dynamics during development and in disease contexts, researchers and large consortia such as the ENCODE Consortium[27], the International Human Epigenome Consortium (IHEC)[28], the NIH Roadmap Epigenomics Mapping Consortium[29] and the BLUEPRINT epigenome project[30], have collected and compared chromatin landscapes across cell types and during disease development.

Over the past decades, we have witnessed the development and widespread use of several chromatin accessibility profiling methods[31–41]. Generally, these methods are based on the physical accessibility to enzymes that fragment, tagment, or methylate DNA in chromatin. Initial screens in the 1970s showed that regions of active transcription were particularly sensitive to digestion by DNA endonucleases, such as deoxyribonuclease I (DNase I), indicating a more permissive form of the chromatin[42], and that chromatin digested at regularly spaced sites due to nucleosome phasing[2,43]. Still today, DNase I is the reagent of choice for TF footprinting, which can determine the location of TF binding sites due to the protection of the site by the TF itself[44–46]. With the advent of next generation sequencing (NGS) techniques, DNase I hypersensitive site sequencing (DNase-seq) was the first adaptation to perform genome-wide profiling of accessible chromatin[32,37]. This was followed by the development of a handful more methods, of which Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and variants[33–35] together with DNase-seq are the two most commonly used chromatin accessibility profiling methods today[47]. As these methods are high-throughput sequencing-based, the analysis of the generated genomics data relies heavily on bioinformatics,

not only for the initial processing but also to biologically interpret chromatin accessibility profiles and to perform more intricate downstream analyses.

Importantly, as regulatory regions co-define a cell type, their accessibility is cell type-dependent, especially for distal regulatory regions[10,48,49]. When investigating heterogeneous samples, it is therefore advisable to measure chromatin accessibility at a single-cell level as bulk methods yield population-averaged accessibility profiles (**Fig. 1**). Currently, the field of single-cell omics, including single-cell epigenomics such as single-cell ATAC-seq, provides exciting new opportunities to study genome regulation in complex tissues such as the brain, whole embryos and tumors[50–57]. Accompanied by the rise of several single-cell chromatin accessibility profiling technologies, a wide range of bioinformatics tools have been developed that allow analysis of the generated data, which is intrinsically sparse[58–67].

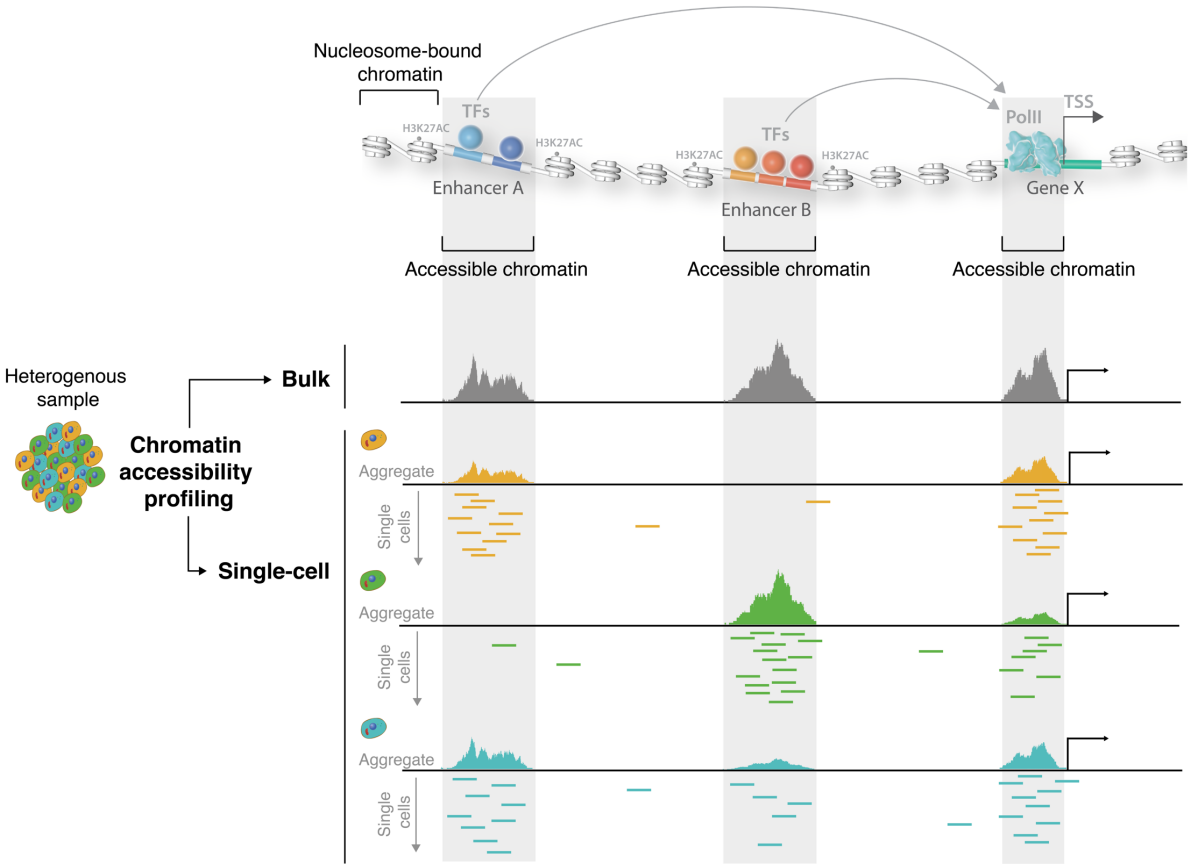

Figure 1. Chromatin accessibility profiling at bulk and single-cell level reveals putative regulatory regions. On top, a representation of a chromatin landscape is shown in which TF-bound enhancers and the promoter of a gene are nucleosome depleted. On the bottom, accessibility profiles of a heterogeneous sample are visualized when measuring chromatin accessibility in bulk or on single-cell level.

Although chromatin accessibility profiling methods may serve as an analytic foundation to identify regulatory regions, it is reported that only around 10-26% of accessible regions in human are active as enhancers[68,69]. Interestingly however, work in both the *Drosophila* embryo[51] and the *Drosophila* eye imaginal disc[70] shows that when a genomic region is uniquely accessible in a specific cell type, the success rate for corresponding enhancer activity is above

80%[71]. In addition, linking (active) accessible regulatory regions to their target genes solely based on accessibility data remains a challenge. Therefore, additional data, including transcriptomics, enhancer-reporter assays, and 3D chromatin architecture maps, especially when combined in a multi-omics fashion, help to determine the function of an accessible region and identify its putative target genes[70,72–77].

This Primer on chromatin accessibility profiling methods provides an overview of commonly used and most recent methods to profile chromatin accessibility, both at bulk and single-cell level. In addition, it provides an outline of computational analysis techniques and examples of applications in diverse organisms and fields. Lastly, the Primer discusses standards for data deposition and examines currently unmet needs and future possibilities to increase our understanding of chromatin accessibility landscapes and their functional role in gene regulation during development, evolution and in disease contexts.

2. Experimentation

2.1. Experimental assays for analyzing bulk cell chromatin accessibility

Chromatin accessibility is traditionally probed by assays such as digestion by nucleases or restriction enzyme digestion, typically at a few selected genomic regions each time[43]. However, NGS has revolutionized the way that chromatin is investigated by allowing us to study its accessibility genome-wide. In the following section, we will briefly describe the principles, pros and cons of several commonly used experimental techniques to assess chromatin accessibility or nucleosome positioning in bulk, including (1) DNase-seq, (2) ATAC-seq, (3) MNase, (4) ChIP-seq, and (5) single-molecule chromatin accessibility profiling methods. Lastly, a variety of less commonly used chromatin accessibility and nucleosome positioning methods are described in **Box 1**.

2.1.1. DNase-seq

The first genome-wide profiling of accessible chromatin was performed in 2008 by sequencing genomic DNA fragments following digestion by DNase I, an endonuclease that preferentially introduces double-stranded breaks in accessible chromatin, a technique referred to as DNase-seq[32,37] (**Fig. 2a**). In a DNase-seq workflow, nuclei are first isolated and permeabilized using a mild detergent such as 0.1% Triton X-100, so that the DNase I enzyme can enter the nucleus efficiently. After digestion, the small DNA fragments (50-100bp) are purified and size-selected for downstream library construction and sequencing. Note, since DNase I digestion is a continuous process, it is necessary to titrate the amount of DNase I to achieve optimal activity when using a new type of cells, or when using DNase I from a different manufacturer or from a different batch. Next to fresh cells, DNase-seq has also been applied fixed (FFPE) samples[32,10]

Major limitations of the traditional DNase-seq include the large number of cells (millions) required as input materials and its tedious and lengthy protocol[78]. In addition, caution must be taken when interpreting DNase-seq results because they show some intrinsic bias in cleavage sites[79,80], which should be considered when interpreting the footprint of a TF[81].

2.1.2. ATAC-seq

ATAC-seq emerged as an alternative assay to investigate accessible chromatin profiles[33]. In this assay, a genetically engineered hyperactive DNA transposase (Tn5) transposes preloaded monovalent mosaic end (ME) adapters to accessible or nucleosome-depleted chromatin regions and tags the DNA with the ME sequence simultaneously[33,82,83] (**Fig. 2b**). The target DNA fragments are purified, PCR-amplified, and sequenced by NGS platforms. Note that sequences detected by ATAC-seq have been found to be highly enriched in DNAse hypersensitivity sites (DHSs)[84–86].

ATAC-seq and its variants[34,35] are sensitive assays that work well on low-input samples (for example 500-50,000 cells) and require a simplified library preparation procedure due to the simultaneous chromatin fragmentation and tagging[33]. In addition to fresh cells and slowly cooled cryopreserved cells, it is possible to generate high signal-to-background profiles from snap-frozen samples using the improved Omni-ATAC protocol[35] or nuclei collected via flow cytometry[87].

Some limitations of ATAC-seq are related to the intrinsic properties of Tn5: (1) it shows steric hindrance and sequence bias in chromatin tagmentation[82,88,89], which would be a challenge for the mapping resolution on both chromatin accessibility and TF footprints. (2) The contamination from organellar DNA, such as mitochondrial DNA and/or chloroplast DNA for plants, or Wolbachia DNA in infected *Drosophila* lines can also increase the sequencing costs as large amounts of sequencing reads can be consumed by these contaminations[33,90]. Organellar contamination can be significantly reduced either by improved lysis condition (as is the case in Omni-ATAC[35]), purification of nuclei via flow cytometry[87] or by applying the clustered regularly interspaced short palindromic repeats (CRISPR) technology to cleave the mitochondrial ribosomal DNA prior to the experiment[91,85]. Another deficiency of the original procedure is that half of all fragments are lost due to the fact that they contain two adapter sequences of the same kind. The Transposome Hypersensitive Sites Sequencing (THS-seq) version of ATAC-seq attempts to rescue the other half of fragments by utilizing a T7 RNA Polymerase linear amplification protocol[92].

2.1.3. MNase-seq

Nucleosome position and occupancy in the genome play key roles in chromatin accessibility. MNase is an endo-exonuclease that cleaves the DNA regions without nucleosome protection and leaves the nucleosome core particles undigested, which can be purified, ligated to adaptors, PCR-amplified and sequenced to reveal genome-wide nucleosome positions (MNase-seq)[39] (**Fig. 2c**).

In MNase-seq, 10,000 to 100,000 of either fresh or formaldehyde crosslinked cells are used for library construction. Digestion of chromatin by MNase typically results in a nucleosome ladder consisting of mononucleosome, dinucleosome, trinucleosome etc., depending on the concentration of MNase in the reaction. The optimal range of digestion usually leads to about 70-80% mononucleosomes and 20-30% higher nucleosome ladders[39].

MNase-seq has been applied to investigate the dynamics of the nucleosome landscape and their function in transcriptional regulation[93]. However, since nucleosome position and occupancy revealed by MNase-seq are based on the average profile of a large number of cells, caution should be taken when interpreting the results, particularly at inactive chromatin regions[94].

2.1.4. ChIP-seq

The N-terminal tails of core histones are enriched with various covalent modifications, which serve as the docking sites for many chromatin-binding proteins[95,96]. Chromatin immunoprecipitation and sequencing (ChIP-seq) is developed to analyze the occupancy of chromatin-binding factors, as well as the landscapes of various histone modifications at a genome-wide level[31,97–99]. Typical histone marks used to define regulatory elements are histone H3 lysine 27 acetylation as this mark correlates well with DNase-seq and ATAC-seq data at TSSs, active promoters and distal active enhancers[100,101]; and H3K4me1 that correlates with poised chromatin states in animals[102]

In ChIP-seq, chromatin is isolated from either formaldehyde fixed cells or non-fixed cells (native chromatin), and fragmented to a range from 100bp to 500bp by sonication or enzymatic digestion[103–105]. Using specific antibodies, the target proteins or histone modifications are captured along with the associated DNA fragments by protein A/G coupled agarose beads or magnetic beads. Then, the DNA fragments are eluted, end-repaired, ligated to adaptors, PCR-amplified and sequenced by NGS.

Traditionally, ChIP-seq needs hundreds of thousands of cells for profiling histone modifications and millions of cells for profiling TFs. The ChIP-seq data quality critically depends on the antibody specificity, the efficiency of chromatin fixation and the residence time of the TF to DNA. Moreover, the whole procedure for ChIP-seq is time-consuming and laborious. In the past decade, several ChIP-seq derivatives have been developed that work with a lower number of cell input and detect TF binding at higher resolution[102,106–109], including a method that combines aspects of ChIP-seq and ATAC-seq (ChIPmentation) by performing tagmentation on immunoprecipitated chromatin fragments[110]. In addition, techniques without the need of prior chromatin fragmentation became available for profiling chromatin modification and TF occupancy on chromatin using hundreds or thousands of cells[111–116], which use antibody guided MNase cleavage or Tn5 tagmentation of chromatin to simplify the procedure of library construction.

2.1.5. Methods based on single-molecule chromatin accessibility profiling

An emerging class of methods aims to map chromatin accessibility and TF binding within single molecules. The advantage of such approaches is that they do not rely on enrichment and provide information about the distribution of accessibility states within the population of chromatin fibers. The assays in this class rely on methyltransferase enzymes that preferentially modify accessible DNA (**Fig. 2D**). For years, the only readout that such methods could rely on was bisulfite conversion of DNA followed by Sanger sequencing (for localized analysis of particular loci)[117–120] and later NGS (for both local and genome-wide coverage), which also

dictated the enzymes used to modify DNA. The first genome-wide assay of this kind was methylation accessibility protocol for individual templates (MAPit[121]), followed by NOMe-seq[122,38], which both use a m5C methyltransferase that modifies cytosines in a GpC context.

As genomes of many eukaryotes contain abundant endogenous methylation in the CpG context, and no non-specific m5C methyltransferases are available with the exception of plants, this is the only modification that can be used to reliably measure accessibility. This presents a serious limitation, as GpC nucleotides are rare in mammalian genomes, only found once every 20 to 30 bp, with much larger stretches of sequence with no informative positions at all being quite common[123]. However, in species such as yeast and *Drosophila*, which lack endogenous methylation, a combination of both a GpC and a CpG methyltransferase can be used, which increases assay resolution down to ~10 bp, a method termed dSMF (digital Single Molecule Footprinting[124]). This approach has proven to be very powerful in enumerating the distinct functional states of individual regulatory elements, down to the ability to footprint the occupancy of individual TFs. Note that the approach could also be applied to mammalian genomes when endogenous methylation is eliminated, although this is not generally applicable as it requires knock out of endogenous methyltransferases[125]. There are additional limitations as only a limited fraction of regulatory regions (typically 30-50%) contains enough informative GpC dinucleotides, and it only provides information about the state of individual molecules within at most 600 bp (the current limit of combined paired-end read length for Illumina sequencing).

The latter issue has been resolved by the advent of long-read sequencing platforms such as PacBio and Oxford Nanopore, which are capable of reading modified bases directly within individual long molecules, though with significantly decreased accuracy compared to bisulfite sequencing. nanoNOMe-seq and MeSMLR-seq (methyltransferase treatment followed by single-molecule long-read sequencing) assays use GpC methylation and nanopore sequencing to map accessibility on a multikilobase scale, though it is still limited in its resolution by available informative positions[126,127].

That limitation has been overcome by taking advantage of the ability of long read platforms to read any modification, and the use of non-specific methyltransferases, such the m6A depositing enzyme EcoGII combined with nanopore or PacBio sequencing, either on total genomic DNA (Single-Molecule long-read Accessible Chromatin mapping sequencing assay (SMAC-seq)[123]; Fiber-seq[128]) or in combination with a phasing MNase digestion step (single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA)[129]). The large number of informative positions allows for fine-scale footprinting almost everywhere in the genome, subject to the limitations imposed by the higher error rate of single-molecule sequencing.

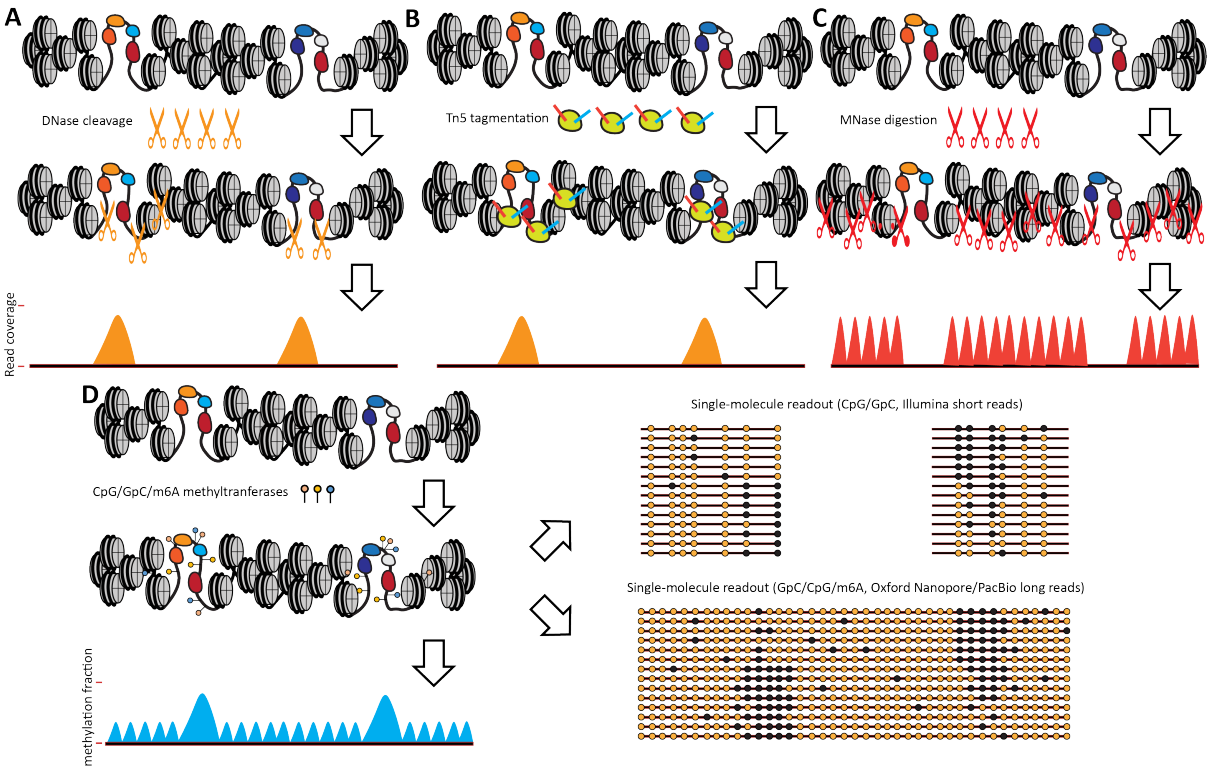

Figure 2. Primary experimental approaches for measuring chromatin accessibility and nucleosome positioning. **A**, In ATAC-seq, hyperactive version of the Tn5 transposase is used to preferentially insert into accessible chromatin while simultaneously attaching adapters to the resulting fragments that can be used to directly amplify sequencing libraries. **B**, In DNase-seq, the DNase I enzyme is used to preferentially cleave accessible chromatin, generating fragments that can then be amplified into sequencing libraries. Both ATAC-seq and DNase-seq generate peaks in read coverage over accessible regions in the genome. **C**, In MNase-seq, the MNAse enzyme is used to digest unprotected DNA, leaving intact fragments protected by protein occupancy (primarily nucleosomes). These fragments are then amplified, resulting in increased read coverage over positioned nucleosomes. **D**, Methyltransferase-based approaches, such as NOMe-seq, dSMF, SMAC-seq, nanoNOMe/MeSLMR-seq and SAMOSA, rely on the labeling of accessible DNA within open chromatin regions and over nucleosome linkers with DNA methylation modifications. These modifications can be m5C methylation in GpC and CpG contexts and also m6A methylation. Bisulfite conversion or the EM-seq assay can be used to convert fragmented DNA into Illumina-compatible libraries, resulting in short-range and sparse-coverage single-molecule footprints. Alternatively, long-read sequencing, which can also directly read m6A methylation and take advantage of its much higher density in the genome, can be used, resulting in multikilobase-scale single-molecule footprints. Methyltransferase-based approaches tend to provide a simultaneous readout of both nucleosome positioning and open chromatin regions, appearing as small "bumps" in the methylated fraction of bases over linker regions and larger peaks over regulatory elements, respectively.

Box 1. Other bulk chromatin accessibility profiling methods

A variety of other methods such as nuclease-accessible site sequencing (NA-seq)[130], restriction endonuclease digestion of chromatin coupled to deep sequencing (RED-seq)[131], quantitative DNA accessibility assay (qDA-seq)[132], and occupancy measurement via restriction enzymes and high-throughput sequencing (ORE-seq)[133] have been used to estimate absolute accessibility levels in yeast and mammalian genomes. Nucleosome positioning has now also been probed

using long-read methods, which allow the mapping of the ends of larger nucleosome arrays rather than the single, di-, or at most trinucleosomes measurable with short reads. Nicking enzyme assisted sequencing (NicE-seq)[134] uses a nicking enzyme to probe accessible DNA. Formaldehyde assisted isolation of regulatory elements (FAIRE-seq)[36,135] is based on the preferential release of accessible chromatin during sonication of crosslinked cells. A technique termed as transposase-mediated analysis of chromatin looping (TrAC-looping), which utilizes Tn5 and a bivalent ME adaptor, also detects genome-wide chromatin accessibility in addition to providing genome-wide chromatin interaction information on regulatory regions[136]. Protect-seq[137] was recently developed to assay the inverse of accessible chromatin, strongly heterochromatinized genomic regions, based on their resistance to nuclease digestion. Differential viral accessibility (DIVA)[138,139] utilizes the preferential viral insertion into accessible DNA to map open chromatin regions. Chromatin accessibility profiling using targeted DamID (CATaDa)[140] labels open chromatin using ectopic expression of the *E. coli* Dam methyltransferase. Next to this, reactive small molecules have also been applied to probe the fine-grained features of accessibility, such as Dimethyl sulfate (DMS) (in DMS-seq[141]) and methidiumpropyl-EDTA (MPE) (in MPE-seq[142]).

Lastly, several chemical approaches for direct mapping of nucleosome positions have been developed. The first such method is based on replacing endogenous histone H4 genes with a H4S47C protein variant. The cysteine in position 47 is located close to the nucleosome center position and can be chemically modified and, using copper and hydrogen peroxide catalysis, used to trigger the cleavage of the DNA backbone close to it[143]. This method was first used to precisely map nucleosome positions in the budding yeast *S. cerevisiae*[144], and more recently in mouse embryonic stem cells[145], though its application is somewhat limited in more complex eukaryotes by the large number of copies of histone genes. A more recent conceptually similar approach relies on the H3Q85C mutation, which generates cleavage at positions close to the nucleosome flanks[146].

2.2. Single-cell chromatin accessibility profiling

Innovation in barcoding and microfluidic strategies have recently enabled high-throughput biochemical profiling of chromatin accessibility at single-cell resolution, including single-cell DNase-seq (scDNase-seq[54]), MNase-seq (scMNase-seq[94]) and ATAC-seq (scATAC-seq[147–151]). Of these protocols, scATAC-seq has emerged as a popular and relatively simple approach to profile chromatin accessibility across hundreds to thousands of individual cells, and we will thus focus on multiple experimental methods of this technique. Current scATAC-seq methods rely on either (droplet) microfluidic or fluorescence cytometrical/plate-based partitioning to uniquely label nuclei in isolation. Procedures characteristic to both flavors of scATAC-seq, as well as consideration for experimental design (**Box 2**), are described below.

2.2.1. Single-cell ATAC-seq

Microfluidics scATAC-seq

Droplet-based single-cell partitioning via microfluidic devices has emerged as a powerful approach for single-cell data generation owing to its reproducibly and relative ease of use. In combination with standard sequencing library reagents and instruments, popular microfluidic approaches for scATAC-seq, such as those commercially available from 10X Genomics

(Chromium Next Gem Single Cell ATAC-seq Library Kit)[151] and BioRad (SureCell ATAC-seq Library Preparation Kit)[150], provide all required reagents necessary to produce scATAC-seq libraries. However, these commercial applications require the acquisition of proprietary robotic sample processing devices (Chromium Controller, 10X Genomics and ddSEQ single-cell isolator, BioRad) that are non-standard in most laboratories.

Droplet microfluidic-based scATAC-seq methods are generally composed of four major steps. First, Tn5 adapter integration is performed on the bulk nuclei suspension, similar to traditional ATAC-seq. Second, transposed nuclei are loaded onto an aqueous channel with PCR reagents and proprietary buffers and mixed with gel-beads containing distinct barcodes. To encapsulate individual nuclei in picolitre reactions with a single gel-bead, the aqueous flow is restricted to channels measuring ~55 uM in width[151]. Droplets are produced by exposing the aqueous flow to a continuous stream of oil. Nuclei droplet loading follows a Poisson distribution and nuclei are thus loaded at low concentrations. Third, barcoded sequences with P5 adapters and tail sequences complementary to Tn5-inserted adapters are released from gel-beads following droplet generation; enabling PCR amplification of accessible chromatin fragments within each droplet in isolation. Finally, the droplet-oil mixture is emulsified, purified with magnetic beads, and subjected to bulk PCR to attach sequencing indices and P7 sequences.

Plate-based scATAC-seq

An alternative to microfluidics approach is to physically separate cells into the wells of plates. Straightforward 96- and 384-well scATAC-seq protocols have been published[148], however, their throughput remains limited by the low number of wells available. The adaptation of scATAC-seq to the ICELL8 Single Cell System (Takara Bio), which has 5,084 nanoliter wells, in the form of μATAC-seq[152], increased the throughput of the assay to a few thousand cells.

Combinatorial indexing (sciATAC-seq)

Higher throughput can be achieved using a combinatorial indexing strategy, as implemented in sciATAC-seq[51,52,149]. In contrast to microfluidic approaches, sciATAC-seq can be performed with access to standard instruments and reagents, with the exception that it requires custom-made Tn5. The core idea behind combinatorial indexing is the repeated pooling-and-splitting of cells or nuclei coupled with labeling of DNA fragments at each step, in such a way that statistically each cell or nucleus is tagged with a unique combination of barcodes. In the simplest implementation of sciATAC-seq, nuclei are distributed into wells containing uniquely indexed Tn5 transposomes, in which tagmentation is performed. Nuclei are then pooled and distributed into the wells of a second plate at numbers sufficiently low to minimize the generation of doublets. The reactions in these wells are then subjected to indexed PCR, generating statistically unique barcode combinations for each cell. Additional rounds of barcoding is also possible, utilizing ligation of barcodes to transposed fragments[153–155], vastly increasing potential throughput. Another approach for increasing throughput is to combine upstream transposition of barcoded Tn5 with a droplet-based scATAC platform such as 10X

or BioRad, in the form of droplet combinatorial indexing, or droplet-based single-cell combinatorial indexing for ATAC-seq (dsciATAC)[150].

Box 2. Experimental design of a scATAC-seq experiment

Similar to other sequencing-based profiling methods, scATAC-seq is susceptible to batch effects that can obscure biological variation. Careful attention to experimental design is central to mitigating batch and other sources of technical variation that strongly depend on the goals of the researcher. For example, in atlas and test versus control studies, a common objective is to contrast regulatory patterns among and within cell-types found in different tissues and organs, or between treatments and control samples. For such cases, scATAC-seq libraries should be constructed in parallel from as many sample types as possible, and should include at least two biological replicates, permitting resources, with the exception of ultra-dense time series or titration series, where the individual wells are intrinsically replicating each other. Prioritizing sample type diversity in preparations from individual batches aids in the mitigation of technical effects and allows researchers to average environmental and genotype influences across replicates. In contrast, comparison of two scATAC-seq libraries produced from separate preparations and from different samples will be confounded by batch effects, resulting in misleading or even erroneous results due to inflated variance between samples. Computational removal of batch effects from single-cell data has been a major focus of many informatics laboratories and shows promise in correcting mistakes stemming from poorly constructed experimental design (see Results). However, there is currently no accepted method to reliably remove all batch effects while preserving biological variation in the absence of true biological replicates. Thus, in cases where generating and sequencing scATAC-seq libraries in different batches is unavoidable, it is pertinent that the researcher takes note of possible sources of variation among samples.

3. Results

In general, a chromatin accessibility analysis workflow consists of three main steps (1) pre-processing, (2) peak calling and (3) downstream analysis (**Fig. 3**). The latter can include differential accessibility analysis, annotation, footprinting, motif enrichment and integration with other omics data. Additional computational steps are needed for single-cell ATAC-seq data. We will discuss each of the steps in more detail and mention commonly used bioinformatics tools. Although there is not yet a gold standard in the field, some general pipelines, such as the ENCODE pipeline for ATAC-seq analysis[156], exist and propose specific tools and a guided workflow for the analysis of chromatin accessibility data.

3.1. Pre-processing

Like most high-throughput sequencing data (**Fig. 3A**), pre-alignment quality control is recommended for chromatin accessibility data and can for instance be performed using *FastQC*[157] or *MultiQC*[158] to examine sequencing quality, GC bias and overrepresented sequences (**Fig. 3B**). Next, sequencing adaptors should be removed using tools like *cutadapt*[159], *trimmonmatic*[160] and *fastq-mcf*[161], which require the input of known Illumina adaptor sequences. Depending on the experimental techniques and when paired-end reads are available,

a size selection can be performed at this point. For instance, removal of multi-nucleosomal reads is advised for MNase-seq data, and for the 'double-hit' DNase-seq protocol an additional *in silico* filtering for fragment inserts between 50-100 bp for TFs binding site detection can be performed in addition to the gel-based or SPRI-based experimental size selection[81,162]. Trimmed and filtered reads are mapped to an organism-specific reference genome. The most widely used aligners for chromatin accessibility data are *Bowtie2*[163] (used in the ENCODE ATAC-seq pipeline[156]), *bwa-mem*[164] (used in the Cell Ranger ATAC Algorithm) or *STAR*[165] (**Fig. 3C**). Following alignment, some additional filtering steps are advised to discard reads with low mapping quality or multi-mapped reads, PCR-duplicated reads, ENCODE blacklisted regions[166] and mitochondrial reads (specifically important for ATAC-seq data in which these can make up as high as 75% of the total amount of mapped reads when using the original protocol[33]) (**Fig. 3D**).

An additional quality control step is recommended at this point by visualizing accumulated read abundance around transcription start sites, which are generally highly accessible[167] (**Fig. 3E**). In addition, visually inspecting the distribution of reads across the genome using genome browsers such as IGV[168], UCSC[169] or JBrowse[170,171] can further increase insight in the quality of the samples (**Fig. 3H**).

3.2. Peak calling

Following initial read processing and quality control comes one of the crucial steps in chromatin accessibility data analysis, namely defining so-called 'peaks' or locations with a high accumulation of reads compared to the background (**Fig. 3F**). These peaks form the basic units in most of the downstream analyses. The most widely used tool for peak calling is *MACS2*[172], which is also the default in the ENCODE ATAC-seq pipeline[156]. *MACS2* is a model-based algorithm originally designed for ChIP-seq data analysis, and implements a dynamic Poisson distribution to capture local background biases in the genome and to effectively detect peaks[172]. Other general (e.g. *ZINBA*[173]) or more technology-specific peak callers exist, e.g. *HMMRATAC*[174] for ATAC-seq; *F-seq*[175] and *Hotspot*[176] for DNase-seq and ATAC-seq. MNase-seq is actually an orthogonal assay compared to the other discussed chromatin accessibility profiling methods as it measures nucleosome-occupied regions. It is therefore the method of choice to map nucleosome positions genome-wide, for which specific tools have been developed[177,178] (**Fig. 3I**). ATAC-seq also lends itself for nucleosome positioning by for instance using the tool *NucleoATAC*[179]. An important parameter to consider during the peak calling step is the signal threshold, which influences the sensitivity and specificity of peak retrieval. The default minimum false discovery rate (FDR) cutoff of 0.05 for *MACS2* has been shown to be optimal for a range of DNase-seq datasets[180].

As datasets often comprise different samples, the construction of a common set of features (i.e., genomic intervals) is crucial in order to be able to compare samples to each other in downstream steps. Usually, a consensus peak file comprising a set of merged peaks across the samples is used. The ENCODE pipeline provides a possible workflow with merge and filter

steps for this objective[156], although other tools can serve the same purpose (e.g. *consensusSeekeR*[181]). Alternatively, a pre-defined set of regions or a binned genome can be used as features in downstream analyses[51,70]. For human and mouse studies, the ENCODE SCREEN regions[182] provide comprehensive sets of intervals, as well as two recently published catalogs of consensus DHS regions (926,535 for human and 339,815 for mouse). For species with more compact genomes and higher regulatory density, such as *Drosophila*, a set of 134,000 regions covering the entire non-coding genome may be used[70].

An important quality control step is to calculate measures that represent the signal-to-noise ratio, which is usually done by calculating the fraction of reads in called peaks (FRiP score). For ATAC-seq the FRiP should preferably be greater than 0.2-0.3 for mammalian species, and the signal proportion of tags (SPOT score) for DNase-seq should exceed 0.4 for mammalian species (i.e. 40% of mapped reads within DHSs)[156,183]. Note that these metrics vary depending on the organism, and can be dependent on the size and complexity of the genome.

Lastly, to ensure reproducibility in the data, ENCODE guidelines recommend that each ATAC-seq experiment should have two or more biological replicates and that replicate concordance should be checked by calculating Irreproducible Discovery Rate (IDR) values[184].

3.3. Downstream analysis

Usually chromatin accessibility profiling is performed on multiple samples, comparing treatment versus control, comparing multiple tissues, or comparing cells during a differentiation process. A central question is to define the set (or signature) of peaks that is differentially accessible in each sample (**Fig. 3G**). For a pairwise comparison between two conditions, differential peak calling can be performed, for example using *MACS2*, in which mapped BAM files of treatment and control samples are provided. Alternatively, statistical analyses can be performed on the read count matrix, with consensus peaks as rows and the different conditions or samples as columns. For pairwise comparisons, several approaches have been borrowed from the RNA-seq field, including MA-plots, and statistical analyses based on the negative binomial distribution, implemented in the *DESeq2*[185] and *edgeR*[186] packages (**Fig. 3G**). Tools like *DiffBind*[187], *HOMER*[188], or *DBChIP*[189] rely on this strategy.

For multi-sample studies, the normalized (e.g., reads per million) region-count matrix can be used for dimensionality reduction and clustering, for example by hierarchical clustering or k-means (**Fig. 3G**). Such clustering algorithms are for instance implemented in the *deepTools* package[190]. The differentially accessible regions can be visualized in a heatmap (**Fig. 3G**). Other researchers have drawn inspiration from tools designed for clustering of regions in single-cell epigenomics data using factor analysis and unsupervised learning. For instance, topic modelling or non-negative matrix factorization, in which a high-dimensional dataset is approximated by a reduced number of representative components, can be applied directly to bulk datasets, or to a matrix with simulated single-cells, created from bulk samples using a bootstrapping procedure[59,191].

To gain biological insight in the sets of cell type specific regions identified via differential accessibility analysis, region set enrichment analysis via *GREAT*[192], *ChIPseeker*[193], *ChIPpeakAnno*[194], *Enrichr*[195], *cisTarget*[196,197], and *LOLA*[198] are used to (1) identify correlations of peaks sets with genome annotation (e.g., promoter, intronic, intergenic) or with existing ChIP-seq tracks; and (2) to couple peaks to the nearest gene, followed by Gene Ontology or pathway enrichment (**Fig. 3L**). In addition, chromatin segmentation approaches such as *ChromHMM*[199], *EpicSeg*[200] and *Segway*[201] are used for genome-wide classification of genomic regions based on epigenomic marks (mostly based on histone modification ChIP-seq) into chromatin states, such as 'active promoter' or 'weak/poised enhancer' per cell type. These annotations can be useful to aid interpretation of gained or lost accessible regions in a study.

As combinatorial binding of TFs to accessible regulatory regions forms the basis of gene regulation, one of the major downstream analysis steps is unravelling which TFs are bound to a set of cell type-specific or differentially accessible regions. Since TFs recognize and bind to TF-specific DNA sequences, we can leverage the enrichment of TF motifs in a set of sequences (**Fig. 3K**). Two major classes of motif analysis tools exist. The first class of tools, e.g. *HOMER*[188], *MEME*[202] and *cisTarget*[196,197], rely on databases of predefined TF motifs (Position Weight Matrices or PWMs[203]), such as *JASPAR*[204], *CIS-BP*[205], *TRANSFAC*[206] and *HOCOMOCO*[207]. These approaches scan the DNA sequences of accessible regions with PWMs, and perform an enrichment analysis compared to a background set or compared to the entire genome as background. The second class, (e.g., *RSAT*[208], *MEME*[202], *Weeder*[209] and *HOMER*[188]) perform *de novo* motif discovery, allowing an unsupervised identification of enriched TF motifs.
Going beyond motif discovery, machine-learning methods have shown promising results, because large sets of co-accessible peaks can be derived per cell type[52]. Examples of convolutional neural network models include *DeepATAC*[210], *DeepLIFT*[211] and *DeepMEL*[212]. Often, these models capture important TF motifs across the training regions but are also able to predict their importance at single-nucleotide resolution within the regulatory sequences. Note that whereas most motif discovery tools require a set of cell type specific peaks, *MEDEA* extracts cell-type-specific peaks from just one input sample using a panel of peaks from reference cell types (e.g. ENCODE-DREAM) prior to a TF motif enrichment analysis[213]. Altogether, motif detection on a set of specifically accessible regulatory regions allows to decode the genome sequences and may reveal possible master regulators that bind to these regions.

An alternative approach to identify TF binding sites from chromatin accessibility data is TF footprinting (**Fig. 3J**). Footprints are small regions (8-30bp) that display relative protection from cleavage due to binding of a TF, and thus correspond to dips in the accessibility peak[44,214,215]. DNase I has been and is still the preferred footprinting reagent. ATAC-seq footprinting has been shown to be less accurate than DNase-seq footprints[216], which might be attributed to the large size of the Tn5 dimer and Tn5-specific cleavage biases that are not accounted for in DNase-seq-designed footprinting algorithms[33,217]. Analytic genomic footprinting approaches either *de novo* annotate DNase I footprints (e.g. the Wellington

algorithm[218], *HINT*[219], *DBFP*[220] and *DNase2TF*[221]); or determine TF occupancy at specific genomic location (e.g. *CENTIPEDE*[222] and footprint likelihood ratio (*FLR*)[223])[224]. Nevertheless, thanks to the success of DNase-seq data for footprinting, footprinting analysis on ATAC-seq data has also been attempted by several groups, for instance in the initial ATAC-seq publication[33], using *DeFCoM*[225] or ATAC-seq-specific footprinting algorithms such as *HINT-ATAC* that consider ATAC-seq artefacts[217]. Note that TF footprinting comes with some limitations as it requires extremely deep sequencing, ideally at least 200 million uniquely mapped reads from a DNase-seq experiment[224], and it is biased by short residence times for some TFs and by intrinsic sequence preferences of DNase I[226].

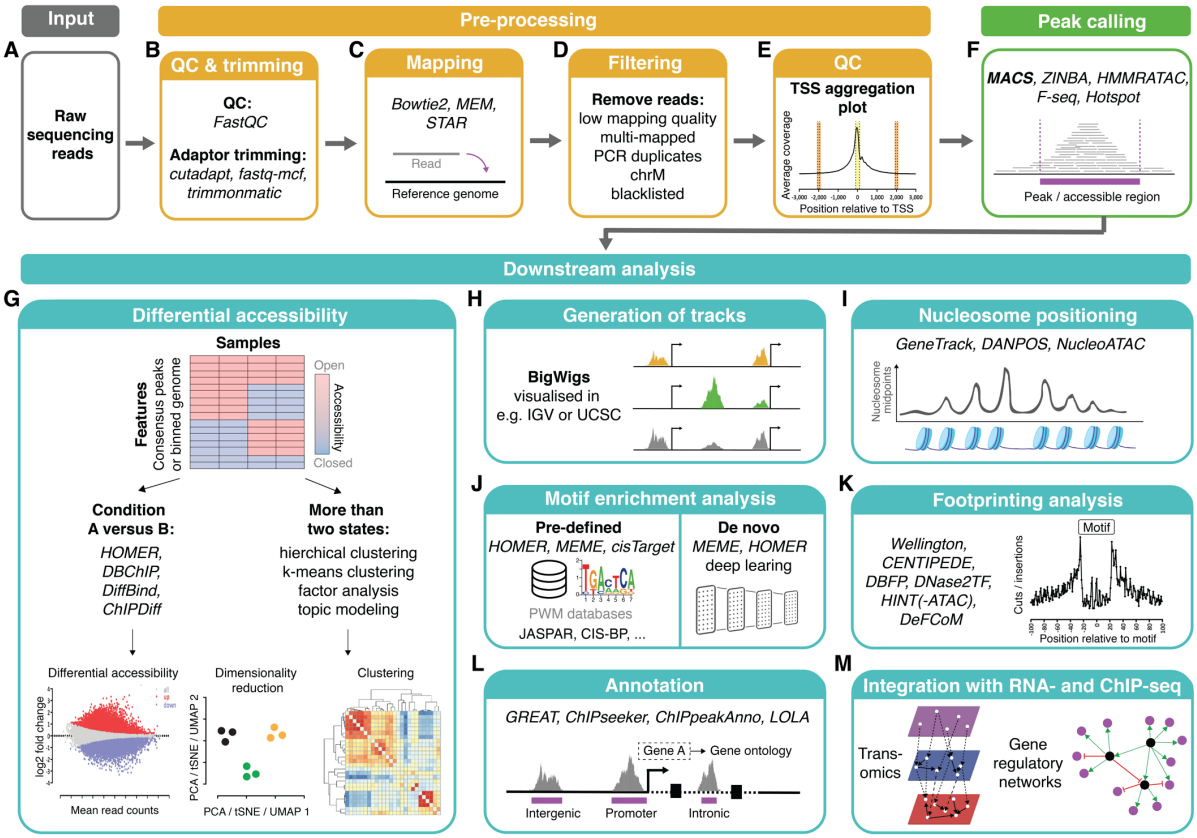

Figure 3. Overview of common bulk chromatin accessibility measurement processing and analysis tasks. Starting from raw sequencing reads (**a**), bulk chromatin accessibility data undergoes pre-processing (**b-e**), followed by peak calling (**f**) and downstream analysis steps (**g-m**).

Single-cell chromatin accessibility data analysis

Single-cell chromatin accessibility data requires similar upstream processing steps as bulk data, including alignment, feature definition and the generation of a count matrix (**Fig. 4A**). However due to the substantial scale and sparsity of the region-by-cell count matrix, specialized bioinformatics tools have been developed, mostly for scATAC-seq data, to handle these assay-specific challenges[58–67]. One major point in which these tools differ is the way they define genomic regions to be used as features [e.g. peaks from bulk or aggregated single-cell data (*chromVar*[65], *Cicero*[64], *cisTopic*[59], *scABC*[67], *Scasat*[58]), pseudo-bulk samples (*Cusanovich et al., 2018*[52]) or fixed size bins (*Cusanovich et al., 2018*[52], *SnapATAC*[61], *ArchR*[62]))] and what the

count features represent [e.g. counting reads in peaks (*cisTopic*[59], *Cusanovich et al., 2018*[52], *scABC*[67], *Scasat*[58]), counting (gapped) k-mers under peaks or around transposase cut sites (*BROCKMAN*[60], *chromVAR*[65]) or counting reads overlapping TF motifs in peaks or genome-wide (*chromVar*[65], *SCRAT*[63])][227]. Important follow-up steps are transformation (e.g. by binarization) and dimensionality reduction of the feature-by-cell matrix to visualize the cells into a 2D- or 3D-space and to perform further downstream analyses such as clustering to uncover the different populations in the sample and their specifically accessible regions (**Fig. 4B,C**). Once cell clusters are obtained, BAM files of all cells belonging to the same clusters can be aggregated to generate pseudobulk tracks to visualize the data (**Fig. 4D**). Recently, 10 computational methods for the analysis of scATAC-seq data have been benchmarked by Chen et al.[227] demonstrating that *SnapATAC*[61], *Cusanovich et al., 2018*[52], and *cisTopic*[59] performed best in distinguishing cell populations in both synthetic and real datasets. Note that compared to scRNA-seq frameworks, there are no designated tools that correct for batch effects in scATAC-seq data, but batch correction is performed inexplicitly during the processing steps such as during feature selection or dimensionality reduction[228]. Batch correction tools designed for scRNA-seq data may be used with precautions to not remove biological variance. Batch effect removal becomes especially important when combining multiple runs into atlases or when integration with scRNA-seq data, for which *BBKNN*[229], *Scanorama*[230] and *scVI*[231] performed best in a recent benchmark[232]. Like in scRNA data, reconstruction of a pseudotime trajectory based on scATAC-seq data can be helpful when studying a system following a cellular differentiation, for instance during embryonic development[233] or hematopoiesis[234] (**Fig. 4F,G**). Tools like *Cicero*[64] (via implementing a modification of the scRNA-seq trajectory inference tools *Monocle*[235]) and *STREAM*[234] have been used to infer such trajectories from scATAC-seq data.

As the complexity of a system or disease exists across all molecular layers, computationally integrating multiple omics modalities holds great promise to achieve a systems biology view and to reconstruct gene regulatory networks. Especially the integration of chromatin accessibility profiles with ChIP-seq and RNA-seq data are of interest (**Fig. 3M**). As TF binding site enrichment within regulatory regions may elude to TF binding, correlation with TF ChIP-seq tracks[197]; or enrichment/overlap of TF ChIP-seq signal/peaks within accessible regions can validate the predicted target sites. For the reconstruction of regulatory networks, specifically the integration of epigenomics and transcriptomics is of interest as this may predict links between accessible regulatory regions and target genes (**Fig. 4E**). An example from the single-cell field is the study by Cao et al.[236] where the authors used a least absolute shrinkage and selection operator (LASSO) model to correlate a gene's expression level with the accessibility of all peaks within 100kB around its TSS , linking 1,260 distal regions to 321 potential target genes, which improved predictions of gene expression based on accessibility profiles by a fourfold as compared to only using chromatin accessibility at promoters.

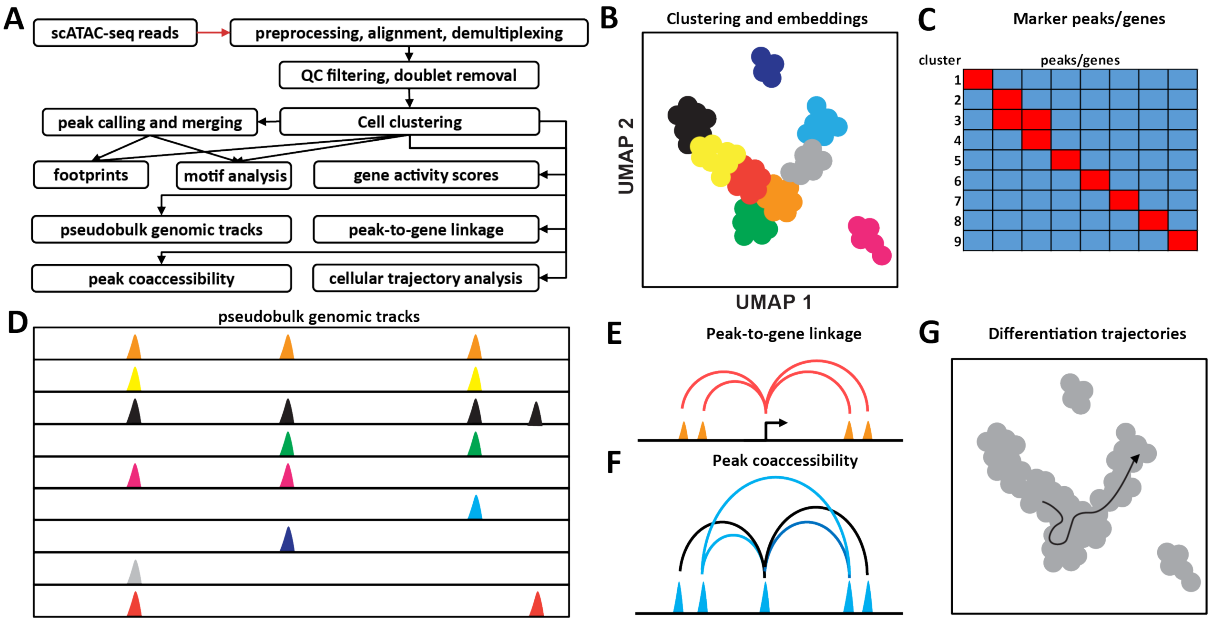

Figure 4. Overview of common scATAC-seq processing and analysis tasks. A, Outline of key steps in processing scATAC-seq datasets. **B**, Clustering of cell types and UMAP embedding of single cells. **C**. Identification of marker genes and/or peaks. **D**, Generation of pseudobulk genome browser tracks for each cell type. **E**, Identification of peak-to-gene links. **F**, Assessing peak coaccessibility. **G**, Differentiation trajectories analysis.

4. Applications

Chromatin accessibility profiling is widely useful for applications in biology and biomedicine, ranging from the analysis of gene regulation and cellular states (section 1 below) over the dissection of healthy and diseased tissues and organs (sections 2 and 3) to the investigation of populations and species (sections 4 and 5). These applications profit from the high genomic resolution of chromatin accessibility profiling and from its relative ease and throughput of these assays.

4.1. Regulation of chromatin accessibility

As nucleosomal occupancy of DNA is refractory to TF binding and transcription, regulation of chromatin accessibility is key to gene regulatory mechanisms. Multiple mechanisms for accomplishing it can be conceived, have been proposed and have some evidence in their support.

Nucleosomes appear to have clear preference for certain sequences, and this bias seems to play some role in establishing nucleosome positions in yeast[237,238], but it is less predictive of nucleosome positioning in metazoan genomes[239,240], and it is mostly not relevant to the major aspect of accessibility regulation, which involves relatively large nucleosome depleted regions associated with active *cis*-regulatory elements. Regulation of regulatory element accessibility and activity is accomplished through the combined dynamics action of TFs, RNA polymerases, chromatin remodeling complexes, histone chaperones and histone variants.

Many TFs only bind to DNA when it is accessible. As a classic example, the vast majority of occupancy sites of glucocorticoid nuclear receptor GR, following its activation by binding by its cognate ligands, are located in pre-existing open chromatin regions[176]; and this property also has direct implications for the cell type-specific effects of its activation.

However, many developmental processes involve the opening up of previously inaccessible chromatin. This process is at its most extreme upon zygotic genome activation (ZGA) during embryonic development, when transcription of the zygotic genome is turned on, but it is also key to all subsequent lineage-specifying developmental transitions, responses to many external and internal stimuli, as well as to cellular reprogramming. A subset of TFs are capable of binding at previously inaccessible chromatin, and subsequently initiating chromatin remodeling leading to an accessible state, and are thus termed "pioneer" factors. Well-known examples of such factors include Zelda, which acts upon ZGA in *Drosophila*[241–243], the Nanog/Oct/Sox pluripotency factors[244–246], FoxA[16], and numerous others[247]. Pioneer factors do not create and maintain an active and accessible state on their own, but this is accomplished together with the recruitment of other TFs, chromatin remodelers and reposition nucleosomes, and chromatin modifiers that deposit histone marks characteristic of active regulatory elements.

What exactly constitutes a pioneer factor and how such TFs exert their action mechanistically has been a subject of much debate, and multiple alternative models have been proposed[1]. Under the strictest definition, a pioneer factor is a TF that directly binds nucleosomal DNA, for which there is direct *in vitro* evidence for a subset of TFs[248,249,19]; however, this is not necessarily the only or even primary mechanism of pioneer action. TFs could be initiating nucleosome displacement through passive competition with core histones for DNA binding during the process of nucleosome turnover[250–252], through binding to linker regions[253,250], or by action in *trans*, i.e. through recruitment of cofactors from an active distal regulatory element[40]. However, what happens *in vivo* is less clear, as no candidate pioneer TFs are known to initiate accessibility at all genomic occurrences of their short degenerate cognate motifs, suggesting a complex context-dependent mechanism of action.

Note that TF binding at regulatory elements in turn can impose constraints on the lateral movement of nucleosomal particles. This is why the most strikingly phased nucleosomal arrays in mammalian genomes are located nearby occupancy sites for strongly and stably bound factors such as CTCF[254] and NRSF[255].

Lastly, cell state transitions also involve the shutting down or "decommissioning" of previously active regulatory elements, which is accomplished by recruiting transcriptional repressors and chromatin modifying complexes removing active chromatin histone marks and depositing repressive ones such as H3K27me3 or H3K9me3, as well as leading to DNA methylation[256].

4.2. Chromatin accessibility across cell types and organs

Chromatin accessibility at gene-regulatory regions is highly dynamic during cellular differentiation and organ development[257,258]. Chromatin accessibility profiling has contributed

to our understanding of chromatin regulation across a broad range of cells in human and mouse[52,183] and in specific organs and cell types. The hematopoietic lineage in particular has served as a blueprint for deciphering the role of chromatin accessibility and epigenetic changes in cellular differentiation[30,259]. Application of ATAC-seq and/or ChIP-seq to FACS-purified hematopoietic cell populations established comprehensive maps of regulatory regions and their dynamic changes in the hematopoietic lineage of human and mouse[34,102,260,261]. Detailed investigations of macrophages connected the regulation of these immune cells to their tissue environment[262,263] while analyses of CD4+ T cells[33,264,265] and innate lymphoid cells[266,267]uncovered a striking degree of plasticity in these immune cell populations. Chromatin regulation in immune cells also contributes to the generation of memory T cells[268] which are poised to implement effector functions upon re-exposure to pathogens, and to the more limited memory of inflammation in regulatory T cells[269] Importantly, immune cell memory is not restricted to B cells and T cells, but also includes monocytes and NK cells[270] and the regulation of such trained immunity appears to involve tightly regulated changes in the epigenomes of the affected cells[271,272].

Beyond the hematopoietic lineage, RNA-seq, ATAC-seq and ChIPmentation profiling of epithelial cells, endothelial cells and fibroblasts from 12 different organs uncovered widespread immune gene regulation in these non-hematopoietic, structural cells, as well as a regulatory potential that appears to pre-program these cells for contributing to pathogen response[273]. Chromatin accessibility has also been studied in neural development[57,274–276] as well as in brain samples of humans[53,55,277] and non-human primates[278]. Notable applications of chromatin accessibility profiling to other cell types and organs have included the analysis of cardiac development[279,280], epidermal progenitor cells in skin[281], and mammary gland development[282]. Finally, initial single-cell atlases of chromatin accessibility across tissues and organs are emerging[52,56], which have the potential to discover new cell types and to define the chromatin states of cell types that are difficult to purify or enrich using FACS. In summary, chromatin accessibility profiling has uncovered a transcription-regulatory landscape that is cell-type-specific and organ-specific, and dynamically changes over the course of cellular differentiation and organ development.

4.3. Chromatin accessibility in human diseases

Changes in chromatin accessibility have been implicated in multiple diseases, where they reflect disease-linked changes in cell composition, gene regulation and epigenetic cell states. Alterations in gene regulation are ubiquitous in cancer and often linked to the developmental abnormalities of cancer cells[283]. In blood cancers, chromatin accessibility patterns have been shown to reflect the cancer's cell-of-origin as well as regulatory changes that appear to contribute to the process of malignant transformation and cancer evolution[34,284–287]. Changes in chromatin accessibility have been investigated over the course of targeted therapy in patients with chronic lymphocytic leukemia[286] and combined with chemosensitivity screening to identify promising drug combination therapies[288]. Chromatin accessibility landscapes have also been mapped in solid tumors, including breast cancer[289], colon cancer[290,291], glioblastoma[292,293],

gastric cancer[294], and lung cancer[295,296]. Pediatric cancers tend to carry particularly pronounced regulatory changes, contrasting with their comparatively low rate of somatic mutations. For example, the *EWS-FLI1* fusion oncogene in Ewing sarcoma has been shown to impose *de novo* enhancers and super-enhancers on the tumor cells[297,298]; and epigenome profiling has uncovered subtype-specific regulatory mechanisms in atypical teratoid rhabdoid tumors[299] and in Langerhans cell histiocytosis[300].

An interesting line of research has investigated the role of the tumor-associated immune cells in solid tumors. Regulatory changes have been implicated in T cell exhaustion in the context of chronic inflammation and the tumor microenvironment[301,302], which compromises the ability of these T cells to fight the tumor. Immunotherapy, most notably blocking of the PD1/PD-L1 pathway, has been shown to revert some of the regulatory changes associated with T cell exhaustion[151,303,304] and is widely useful for the treatment of those solid tumors that have a high degree of immunogenicity[305]. However, not all exhausted T cells can be rejuvenated by immune checkpoint blockade, as some T cells appear to transition to a fixed regulatory state that renders them resistant to reprogramming[301]. In addition to immunotherapy, selective reprogramming of DNA methylation can be used to alter the T-cell landscape resulting in enhanced treatment efficiency[84,306].

Beyond cancer, where chromatin accessibility has been studied most extensively, changes in chromatin accessibility have also been observed in immune diseases such as inflammatory bowel disease[307] and rheumatoid arthritis[308]. Moreover, changes in epigenome and chromatin accessibility profiles have been observed in post-mortem brain tissue from patients with Alzheimer's disease[309], schizophrenia[310] and autism spectrum disorder[311]. In summary, chromatin accessibility profiling of primary patient samples is already widely used for identifying disease-linked changes in chromatin structure and transcription regulation, and there is substantial scope for new discoveries as researchers move beyond cancer and are investigating regulatory mechanisms in many diseases that have yet received little attention.

4.4. Chromatin accessibility variation within populations

Extension of chromatin accessibility assays to populations of diverse genetic backgrounds has proven valuable for advancing our understanding of how sequence variation impacts *cis*-regulation within a species. A striking 90% of disease-associated variants in humans identified via GWAS localize to gene-distal non-coding loci, obfuscating functional predictions[24,312,313]. Mounting evidence has implicated alteration of gene regulation as a key driver of phenotypic evolution and disease proliferation. Quantitative trait loci (QTL) mapping of molecular traits, such as expression variation (eQTL), provides an attractive approach for deciphering the gene regulatory potential of genetic variants within a population. Leveraging a molecular QTL framework, a large-scale DNase-seq panel of 70 lymphoblastoid cell lines from the Yoruba HapMap showed that approximately 50% of chromatin accessibility associated variants coincide with variants associated with expression variation, with the allele conferring increased accessibility generally associated with increased gene expression[314]. This study also provided

evidence that sequence alterations underlying *cis*-elements perturb TF binding affinities, leading to weakened or ablated binding. An analysis of CD4+ T cell chromatin accessibility from 105 healthy donors revealed that only 15% of genetic variants embedded within accessible chromatin regions affect the relative accessibility of the cognate locus[315]. Thus, the majority of genetic variants located within accessible chromatin appear to lack functional consequences. The same study further demonstrated that pairwise correlations of accessible regions (co-accessible regions) readily recapitulates three-dimensional higher-order chromatin interactions as defined by *in situ* HiC data, suggesting that local chromatin accessibility among pairs of regions are coordinated with higher-order genome structure, particularly within the same topologically-associated domains (TADs). In line with these findings, local chromatin accessibility in a subset of regions were associated with variants located 10s to 100s of kilobases away, reflecting putative interactions. Importantly, integration of population-scale accessibility data captured 10-30% of previously reported autoimmune-associated variants and explained 1-7% of disease heritability. In model organisms, chromatin accessibility can be performed across a cohort of homozygously inbred individuals, making the identification of chromatin accessibility QTL (caQTL) more straightward. Jacobs et al., revealed that a critical subset of caQTLs could be explained by making or creating binding motifs for pioneer factors[316]. In an alternative approach, chromatin accessibility can also be compared between alleles, within the same individual, to identify allele-specific chromatin accessibility[317].

Taken together, population-based and/or allele-specific analysis of chromatin accessibility provides a powerful approach for dissecting the regulatory potential of genetic variants associated with a trait of interest. Additional studies in other tissues and disease states leveraging single-cell technologies have the potential to systematically map all chromatin accessibility modifying variants in a cell-type specific fashion.

4.5. Evolution of chromatin accessibility

The use of chromatin accessibility data has greatly facilitated the identification of causal genetic variants underlying disease and trait variation; however, it is also proving useful to study the evolution of gene regulation and morphological evolution between species. For example, major morphological transitions, such as the loss of limbs in snakes and eye degeneration in subterranean mammals, have been linked to loss of regulatory elements[318]. These regulatory regions were discovered using a combination of tissue-specific ATAC-seq and comparative genomics. In another study, chromatin accessibility data in combination with H3K27ac and H3K4me3 was used to identify promoters and enhancers in liver tissue of 20 mammalian species[261]. It was determined that the rate of sequence variation is much greater for enhancers in comparison to promoters. This was reflected by a lower conservation of enhancers between species, yet, newly evolved enhancers were more likely to be under positive selection in a lineage specific manner.

A major advantage of incorporating chromatin accessibility data into these studies is that DNA sequence variation is often too high in intergenic regions to identify *cis*-regulatory elements

using sequence-based alignments alone[319]. This is especially problematic for studies in plants, where sequence turnover between related plant species is much greater than what is observed between related animal species[320]. As such, comparative epigenomics is revealing important clues about the evolution of gene regulation. For instance, rapid evolution of *cis*-regulatory regions has been identified in a comparative epigenomics study of numerous flowering plant species ranging in genome size from ~150 Mb to ~5,000 Mb[321]. The frequency of distal accessible chromatin regions was correlated with genome size and their distal location from genes was mostly likely due to transposon and repeat expansion in these plants[322,75,318].

Lastly, the lack of distal regulatory regions in *Capsaspora owczarzaki*, a unicellular organism sister to other animal species, has led to the hypothesis that distal regulation is a feature of animal multicellularity[323], however, with the increase in profiles of chromatin accessibility across taxa it seems more likely that distal regulation is a consequence of genome size[321]. Additional comparative epigenomic studies of chromatin accessibility across diverse taxa and of species that represent key nodes in the tree-of-life will further unveil diverse mechanisms in the evolution of gene regulatory mechanisms.

5. Reproducibility and Resources

The genomics community has been leading the way in creating standards for data information, data quality and data deposition for decades. This reflects that many genome-wide datasets serve as community resources and, as a result, they are repeatedly used and incorporated into future studies by individual labs. To increase the usability of epigenomics data, it is common practice to submit the data to well-funded and stable data archive facilities such as the Gene Expression Omnibus (GEO) repository[324] at the National Center for Biotechnology Information (NCBI) or to the ArrayExpress database[325] at the European Bioinformatics Institute (EBI). These databases host records of genomics data containing not only count matrices and other useful processed output files (e.g. bigWig files or BED files enriched for chromatin modification or accessibility), but also a short description of the experimental design and processing steps to reach the submitted output files, as well as a link to the archived raw sequencing data. For non-human species and open-consent human donors, the raw sequencing data should be submitted to for instance Sequence Read Archive (SRA)[326], European Nucleotide Archive (ENA)[327] or DNA Data Bank of Japan (DDBJ)[328]. For human donors where controlled access is required, the raw sequencing data should be submitted for instance into the European Genome-phenome Archive (EGA)[329] or the database of Genotypes and Phenotypes (dbGaP)[330] from NCBI. Although rarely required by journals, many researchers are in addition hosting their data in track hubs through publicly accessible genome browsers, as well as other interactive web-based tools to for instance visualise dimensionality reduction plots of scATAC-seq data using SCope[331], a Shiny app[63] or ASAP[332]. This increases data dissemination and provides a user-friendly tool for scientists not as familiar with computational methods for analyzing data.

To facilitate interpretation and reproducibility, the deposited data should include metadata. For example, data entry requirements that are useful to addresses issues associated with

reproducibility could include sources of possible biological variation (i.e. genotype, sex of samples, age, tissue/organ/cell type) and technical variation (i.e. antibodies – lot number, nucleases/integrases – lot number, sequencing library procedure, instrument used for sequencing and type of sequencing run). They are also important variables that can be incorporated into data analyses as covariates or to correct for batch effects. Genome assembly and genome annotation versions used in data analyses should also be provided.

Lastly, distribution of custom code and descriptions of computational methods are also paramount to reproducibility. As one example mentioned above, the ENCODE Consortia has developed extensive open source software that is accompanied with 'best practices' and descriptive details on the rationale for data processing steps, thresholds and quality metrics for data evaluation. In general, software used for data analyses should include the software version and parameter options applied. Custom code should be disseminated through public hosts such as GitHub, or can be archived in a static digital repository such as Zenodo, or on more specialized repositories such as Kipoi[333] for ready-to-use trained machine learning models for genomics. Efforts to address the biological, experimental and computational variables described above will increase reproducibility in addition to the usability of these data for years to come.

6. Limitations and optimizations

While chromatin accessibility has proven a powerful and informative window into gene regulation, accessibility alone must often be linked to orthogonal measurements or perturbations to build a causal or mechanistic understanding of genomic function. While accessibility dynamics can be readily mapped, the specific molecular factors that drive accessibility changes may only be inferred by changes in the accessibility or footprints associated with DNA motifs. However, specific DNA motifs may often be bound by a variety of related protein factors, often within a family of structurally similar DNA binding domains. While motif-specific accessibility changes may be linked to specific TFs based on concomitant changes in gene expression of a specific member of a family, a mechanistic linkage to specific binding requires subsequent experiment, such as expression knockdown or ChIP-seq targeting the specific TF implicated.

Additionally, the accessibility of a putative regulatory locus is likely a necessary but not sufficient criterion for *bona fide* functional regulation. Other markings, such as H3K27ac or the presence of nascent transcription of enhancer RNA appear to mark a subset of accessible elements that are more highly enriched for function[334–336]. Therefore, chromatin accessibility data might be merged with a variety of other genomic assays of function to build a more thickly constituted set of inferences supporting functionality of specific elements.

Finally, many chromatin accessibility methods, notably DNAse-seq and to perhaps a lesser extent ATAC-seq, may require optimization of reaction time, lysis protocols, cell handling, freezing or thawing, as well as library purification, to produce optimal data. For methods such as ATAC-seq, a number of quality metrics exist prior to sequencing, such as relative PCR

cycles required to amplify the library, or the periodicity of the length distribution of fragments generated by the transposition reaction, which allow for relatively rapid and inexpensive optimization of sequencing libraries.

7. Outlook

The past decade has seen an explosion in studies examining chromatin accessibility and its variation in different cell types, tissues, organs and organisms. The current and future challenge is to dissect the function of these regulatory regions in relation to other regulatory layers and gene expression (**Fig. 5**). Chromatin accessibility alone does not reveal the activity state or the functional properties of the region (whether it acts as a promoter, enhancer, silencer), or which factors are bound to the region or its potential role in other functions such as 3D genome topology or replication origins. Moreover, information on the identity of the target genes, and whether a regulatory region is functionally required for gene expression, is also missing.

Many of these challenges can be overcome by a more holistic multi-omics approach, by profiling multiple molecular layers from the same sample, such as the transcriptome, chromatin modifications and TF occupancy, in addition to chromatin accessibility. A common approach is to run multiple omics methods on fractions of the same sample, using protocols optimized separately for each assay, thus generating comparable datasets[337,338]. However, running separate assays can introduce batch effects that are difficult to mitigate computationally, which can be a drawback of this strategy.

Chromatin accessibility profiling in single cells has surged dramatically in recent years, in part due to combinatorial indexing (sciATAC-seq)[149] and the recent availability of commercial kits for droplet-based scATAC-seq[150,151]. We expect further improvements to the assay in the coming years as this trend keeps increasing. In contrast to RNA, which has a high dynamic range, there are only two loci that can be measured simultaneously in a diploid genome by single-cell regulatory genomics-based methods. As a result, the data is mostly binary and still very sparse due to the low coverage per cell, making the analysis of accessibility and other regulatory features at the single-cell level extremely challenging and a certain degree of data aggregation across cells or features is usually required. It is also difficult to estimate the sensitivity of scATAC-seq. Roughly ~10-15% of known peaks are recovered per single cell (PMID: 26083756), but it is actually not known how many regulatory elements are accessible in any given cell at any instance in time. Technical improvements during the past couple of years have boosted cell coverage, which ameliorated both issues to some extent and resulted in a significant increase in data resolution, allowing a sharper distinction between cell types as well as regulatory changes. Given this inherent difference between scRNA-seq and scATAC-seq (and scChIP) data, specialized computational tools have been developed that address the sparsity and binary nature of scATAC-seq data and facilitate more integrated analyses across groups of cells[58–67]. However, the availability of tools designed for scATAC-seq is still very limited when it comes to specific analysis tasks, such as pseudotime and trajectory inference. While comparisons of performance and applicability of scATAC-seq methods have been performed[227], there are no uniform pipelines adapted widely by the community, which

complicates a systematic comparison and interpretation of results coming from different labs. We foresee major efforts in the coming years towards standardizing comprehensive computational pipelines for analysis.

Recent advances in single-cell methods are pushing technologies to perform multi-omic measurements from the same cell. Multiple methods have been published recently for simultaneous single-cell ATAC-seq and transcriptome profiling. These include sci-CAR[236], Paired-seq[155], and SHARE-seq[153], which are all based on combinatorial indexing, as well as droplet-based methods, such as SNARE-seq[339]. Also joint profiling of chromatin accessibility with either protein levels (Pi-ATAC[340]); or with DNA methylation (scNOMe-seq[341]; chromatin overall omic-scale landscape sequencing (COOL-seq)[342], EpiMethylTag[343], methyl-ATAC-seq[344], ATAC-Me[345]); or with a combination of both DNA methylation and transcriptome measurements (single-cell nucleosome, methylation and transcription sequencing (scNMT-seq)[346]) has been achieved.

Several technical challenges have so far limited the widespread application of these methods. Sample fixation, reaction conditions and other experimental parameters are often not compatible for multiple assays, complicating the optimization of joint protocols. Moreover, the resulting data is limited by the combined sensitivity of the methods, for example running two assays each having a 10% capture rate could result in a very small set of overlapping features. Profiling multiple molecular layers raises the non-trivial computational challenge of integrating the datasets. Methods that can handle the harmonization of bulk and single-cell multi-omic measurements have recently been developed (MOFA[347], Seurat v3[66]). A key feature required for future computational methods is flexibility; methods need to handle datasets coming from very different modalities, coming from the same cell or from the same sample and will need to impute missing molecular layers based on the ones that were profiled. Measuring multiple parameters from the same single cell should greatly advance our ability to link regulatory properties and deconstruct regulatory connections. Having information on coordinated changes in distal open chromatin regions (putative enhancers) and gene transcription from the same cell, for example, would greatly help to link enhancers to their potential target genes. We anticipate important developments in both experimental and computational multi-omic approaches in the coming years.

Functionality of accessible chromatin regions can also be probed by perturbation, for example by mutation of key transcription factors. The high degree of cellular heterogeneity in complex systems, such as developing embryos, has limited the usefulness of this approach. However, single-cell accessibility profiling could solve this issue, by identifying the impact of the mutations directly in the affected cell types, revealing both changes in regulation as well as alterations in cell fate decisions. Large-scale perturbation and profiling of regulatory networks has been performed in cell culture models by coupling CRISPR screening with scATAC-seq readout (Perturb-ATAC[348]). In more complex systems, where high-throughput targeted mutagenesis is not feasible, natural sequence variation could be exploited as a large-scale perturbation tool. In this context, profiling accessibility both intra- and inter- species can give insights into regulatory variation and functionality, as discussed above.

Finally, a particularly exciting area of future development is the integration of chromatin accessibility profiling with imaging-based approaches. Current chromatin accessibility profiling protocols involve tissue dissociation to extract cells or nuclei, which leads to the loss of the native spatial context. ATAC-see[349] mitigates this problem by performing the Tn5 reaction *in situ* on microscopy slides and using fluorescent adaptors that are compatible with both imaging and sequencing; and sciMAP-ATAC[350] provides a medium-level spatial mapping of single-cell chromatin accessibility profiles by taking microbiopsies of a tissue prior to the sciATAC-seq workflow. Further integration of ATAC-seq with high-throughput fluorescence in situ hybridization (FISH) and other imaging-based methods will lead to new ways of interrogating the genome of complex systems in situ after stimuli and perturbations.

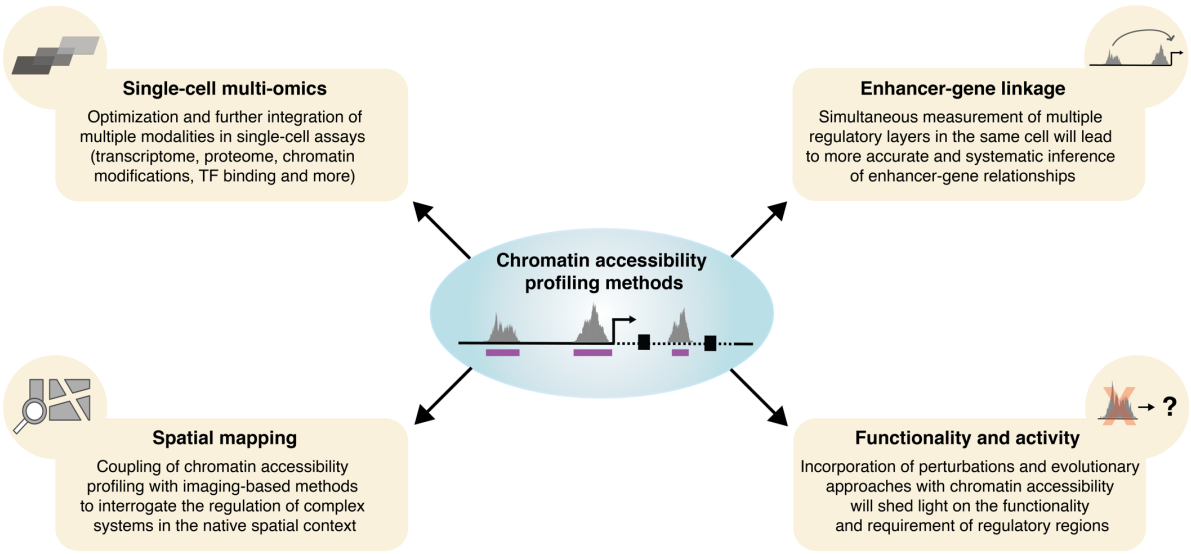

Figure 5. Schematic overview of future roads and opportunities for chromatin accessibility profiling. In the coming years, our capability of measuring chromatin accessibility concurrently with multiple regulatory layers in the same single cell will continue to expand. New insights into regulatory biology will be gained by applying these methods in the native spatial context and in systems undergoing perturbations. Development of computational tools that can dive into the complexity of the emerging datasets will be crucial for the success of these endeavors. Ultimately, these approaches will empower us to functionally dissect the role of regulatory elements and their relationship to gene expression.

Acknowledgements

This work was supported by the Division of Intramural Research, National Heart, Lung and Blood Institute, National Institutes of Health, USA (K.Z.), the National Science Foundation (NSF, IOS-1856627) (R.J.S). R.J.S. is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts. L.M. was supported by a PhD fellowship from the FWO (no. 1S03317N), A.P.M. by an NSF Postdoctoral Fellowship in Biology (NSF, DBI-1905869), C.B. by an ERC Starting Grant (European Union's Horizon 2020 research and innovation program, grant agreement no. 679146), E.F. by an ERC Advanced grant (DeCRypT) and S.A. by an ERC Consolidator Grant (cis_CONTROL).

Competing interests

C.B. is an inventor on a patent describing the ChIPmentation assay, which has been licensed to Diagenode s.a. (Liège, Belgium) and commercialized as a kit and service. R.J.S. is a co-founder of REquest Genomics, LLC, a company that provides epigenomics services. All other authors declare no competing interests.

References

1. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
2. Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **184**, 868–871 (1974).
3. Mazia, D. ENZYME STUDIES ON CHROMOSOMES. *Cold Spring Harb. Symp. Quant. Biol.* **9**, 40–46 (1941).
4. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
5. Woodcock, C. L., Safer, J. P. & Stanchfield, J. E. Structural repeating units in chromatin. I. Evidence for their general occurrence. *Exp. Cell Res.* **97**, 101–110 (1976).
6. Lee, C.-K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900–905 (2004).
7. Ozsolak, F., Song, J. S., Liu, X. S. & Fisher, D. E. High-throughput mapping of the chromatin structure of human promoters. *Nat. Biotechnol.* **25**, 244–248 (2007).
8. Sheffield, N. C. & Furey, T. S. Identifying and characterizing regulatory sequences in the human genome with chromatin accessibility assays. *Genes* **3**, 651–670 (2012).
9. The Mouse ENCODE Consortium *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
10. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
11. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465–476 (2008).
12. Turner, B. M. Defining an epigenetic code. *Nat. Cell Biol.* **9**, 2–6 (2007).
13. Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. Nucleosomes Unfold Completely at a Transcriptionally Active Promoter. *Mol. Cell* **11**, 1587–1598 (2003).
14. Reinke, H. & Hörz, W. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol. Cell* **11**, 1599–1607 (2003).
15. Chaya, D., Hayamizu, T., Bustin, M. & Zaret, K. S. Transcription factor FoxA (HNF3) on a nucleosome at an enhancer complex in liver chromatin. *J. Biol. Chem.* **276**, 44385–44389 (2001).
16. Cirillo, L. A. & Zaret, K. S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol. Cell* **4**, 961–969 (1999).
17. Sherwood, R. I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
18. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
19. Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
20. Hendrich, B. & Bickmore, W. Human diseases with underlying defects in chromatin structure and modification. *Hum. Mol. Genet.* **10**, 2233–2242 (2001).

21. Matsumoto, L. *et al.* CpG demethylation enhances alpha-synuclein expression and affects the pathogenesis of Parkinson's disease. *PloS One* **5**, e15522 (2010).
22. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).
23. Vinagre, J. *et al.* Frequency of TERT promoter mutations in human cancers. *Nat. Commun.* **4**, 2185 (2013).
24. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
25. Moore, S. P. G., Kruchten, J., Toomire, K. J. & Strauss, P. R. Transcription Factors and DNA Repair Enzymes Compete for Damaged Promoter Sites. *J. Biol. Chem.* **291**, 5452–5460 (2016).
26. Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).
27. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
28. Stunnenberg, H. G. *et al.* The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1145–1149 (2016).
29. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
30. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30**, 224–226 (2012).
31. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).
32. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
33. 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**, (2013).
34. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
35. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, (2017).
36. 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).
37. Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
38. Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).
39. Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
40. Taberlay, P. C. *et al.* Polycomb-Repressed Genes Have Permissive Enhancers that Initiate Reprogramming. *Cell* **147**, 1283–1294 (2011).
41. Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R. & Elgin, S. C. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell* **16**, 797–806 (1979).
42. Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).
43. Hewish, D. R. & Burgoyne, L. A. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* **52**, 504–510 (1973).
44. 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).
45. Kemper, B., Jackson, P. D. & Felsenfeld, G. Protein-binding sites within the 5' DNase I-hypersensitive region of the chicken alpha D-globin gene. *Mol. Cell. Biol.* **7**, 2059–2069 (1987).

46. Vierstra, J. *et al.* Global reference mapping of human transcription factor footprints. *Nature* **583**, 729–736 (2020).
47. 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).
48. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
49. West, J. A. *et al.* Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat. Commun.* **5**, 4719 (2014).
50. Al-Ali, R. *et al.* Single-nucleus chromatin accessibility reveals intratumoral epigenetic heterogeneity in IDH1 mutant gliomas. *Acta Neuropathol. Commun.* **7**, 201 (2019).
51. Cusanovich, D. A. *et al.* The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018).
52. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324.e18 (2018).
53. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28**, 1243–1252 (2018).
54. Jin, W. *et al.* Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* (2015) doi:10.1038/nature15740.
55. Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2017).
56. Pijuan-Sala, B. *et al.* Single-cell chromatin accessibility maps reveal regulatory programs driving early mouse organogenesis. *Nat. Cell Biol.* **22**, 487–497 (2020).
57. Preissl, S. *et al.* Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* (2018) doi:10.1038/s41593-018-0079-3.
58. Baker, S. M., Rogerson, C., Hayes, A. & Sharrocks, A. D. Classifying cells with Scasat - a tool to analyse single-cell ATAC-seq. 1–16 (2017).
59. Bravo González-Blas, C. *et al.* cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400 (2019).
60. de Boer, C. G. & Regev, A. BROCKMAN: deciphering variance in epigenomic regulators by k-mer factorization. *BMC Bioinformatics* **19**, (2018).
61. Fang, R. *et al. Fast and Accurate Clustering of Single Cell Epigenomes Reveals* Cis *-Regulatory Elements in Rare Cell Types*. http://biorxiv.org/lookup/doi/10.1101/615179 (2019) doi:10.1101/615179.
62. Granja, J. M. *et al. ArchR: An integrative and scalable software package for single-cell chromatin accessibility analysis*. http://biorxiv.org/lookup/doi/10.1101/2020.04.28.066498 (2020) doi:10.1101/2020.04.28.066498.
63. Ji, Z., Zhou, W. & Ji, H. Single-cell regulome data analysis by SCRAT. *Bioinformatics* **33**, 2930–2932 (2017).
64. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71**, 858-871.e8 (2018).
65. 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**, (2017).
66. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
67. Zamanighomi, M. *et al.* Unsupervised clustering and epigenetic classification of single cells. *Nat. Commun.* **9**, (2018).
68. Kwasnieski, J. C., Fiore, C., Chaudhari, H. G. & Cohen, B. A. High-throughput functional testing of ENCODE segmentation predictions. *Genome Res.* **24**, 1595–1602 (2014).
69. Wang, X. *et al.* High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat. Commun.* **9**, 5380 (2018).
70. Bravo González-Blas, C. *et al.* Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics. *Mol. Syst. Biol.* **16**, e9438 (2020).

71. Graybuck, L. T. *et al. Enhancer viruses and a transgenic platform for combinatorial cell subclass-specific labeling*. http://biorxiv.org/lookup/doi/10.1101/525014 (2019) doi:10.1101/525014.
72. Hafez, D. *et al.* McEnhancer: predicting gene expression via semi-supervised assignment of enhancers to target genes. *Genome Biol.* **18**, 199 (2017).
73. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
74. Moore, J. E., Pratt, H. E., Purcaro, M. J. & Weng, Z. A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. *Genome Biol.* **21**, 17 (2020).
75. Ricci, W. A. *et al.* Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **5**, 1237–1249 (2019).
76. Ron, G., Globerson, Y., Moran, D. & Kaplan, T. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nat. Commun.* **8**, 2237 (2017).
77. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
78. 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.* **5**, 1–12 (2010).
79. Lazarovici, A. *et al.* Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 6376–6381 (2013).
80. Suck, D., Lahm, A. & Oefner, C. Structure refined to 2A of a nicked DNA octanucleotide complex with DNase I. *Nature* **332**, 464–468 (1988).
81. 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).
82. 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).
83. Goryshin, I. Y. & Reznikoff, W. S. Tn5 in vitro transposition. *J. Biol. Chem.* **273**, 7367–7374 (1998).
84. Qu, K. *et al.* Chromatin Accessibility Landscape of Cutaneous T Cell Lymphoma and Dynamic Response to HDAC Inhibitors. *Cancer Cell* **32**, 27-41.e4 (2017).
85. Wu, J. *et al.* The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
86. Wu, J. *et al.* Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).
87. Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M. & Schmitz, R. J. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res.* **45**, e41 (2017).
88. Meyer, C. A. & Liu, X. S. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat. Rev. Genet.* **15**, 709–721 (2014).
89. Sato, S. *et al.* Biochemical analysis of nucleosome targeting by Tn5 transposase. *Open Biol.* **9**, 190116 (2019).
90. Davie, K. *et al.* Discovery of Transcription Factors and Regulatory Regions Driving In Vivo Tumor Development by ATAC-seq and FAIRE-seq Open Chromatin Profiling. *PLoS Genet.* **11**, 1–24 (2015).
91. Montefiori, L. *et al.* Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci. Rep.* **7**, 2451 (2017).
92. Sos, B. C. *et al.* Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. *Genome Biol.* **17**, 20 (2016).
93. 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).
94. Lai, B. *et al.* Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* **562**, 281–285 (2018).

95. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
96. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
97. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
98. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
99. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
100. Bysani, M. *et al.* ATAC-seq reveals alterations in open chromatin in pancreatic islets from subjects with type 2 diabetes. *Sci. Rep.* **9**, 7785 (2019).
101. Shu, W., Chen, H., Bo, X. & Wang, S. Genome-wide analysis of the relationships between DNaseI HS, histone modifications and gene expression reveals distinct modes of chromatin domains. *Nucleic Acids Res.* **39**, 7428–7443 (2011).
102. Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
103. Kuo, M. H. & Allis, C. D. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods San Diego Calif* **19**, 425–433 (1999).
104. O'Neill, L. P. & Turner, B. M. Immunoprecipitation of native chromatin: NChIP. *Methods San Diego Calif* **31**, 76–82 (2003).
105. Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **25**, 99–104 (2000).
106. Brind'Amour, J. *et al.* An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* **6**, 6033 (2015).
107. Dahl, J. A. *et al.* Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).
108. Ng, J.-H. *et al.* In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev. Cell* **24**, 324–333 (2013).
109. Zhang, B. *et al.* Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
110. 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).
111. 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).
112. Harada, A. *et al.* A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.* **21**, 287–296 (2019).
113. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
114. Ku, W. L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat. Methods* **16**, 323–325 (2019).
115. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, (2017).
116. Wang, Q. *et al.* CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Mol. Cell* **76**, 206-216.e7 (2019).
117. Carvin, C. D., Dhasarathy, A., Friesenhahn, L. B., Jessen, W. J. & Kladde, M. P. Targeted cytosine methylation for in vivo detection of protein-DNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7743–7748 (2003).
118. Jessen, W. J. *et al.* Mapping chromatin structure in vivo using DNA methyltransferases. *Methods San Diego Calif* **33**, 68–80 (2004).
119. Kladde, M. P., Xu, M. & Simpson, R. T. Direct study of DNA-protein interactions in repressed and active chromatin in living cells. *EMBO J.* **15**, 6290–6300 (1996).

120. Xu, M., Kladde, M. P., Van Etten, J. L. & Simpson, R. T. Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. *Nucleic Acids Res.* **26**, 3961–3966 (1998).
121. Pardo, C. E., Nabilsi, N. H., Darst, R. P. & Kladde, M. P. Integrated DNA methylation and chromatin structural analysis at single-molecule resolution. *Methods Mol. Biol. Clifton NJ* **1288**, 123–141 (2015).
122. Darst, R. P., Nabilsi, N. H., Pardo, C. E., Riva, A. & Kladde, M. P. DNA methyltransferase accessibility protocol for individual templates by deep sequencing. *Methods Enzymol.* **513**, 185–204 (2012).
123. Shipony, Z. *et al.* Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).
124. Krebs, A. R. *et al.* Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Mol. Cell* **67**, 411-422.e4 (2017).
125. Sönmezer, C. *et al. Single molecule occupancy patterns of transcription factors reveal determinants of cooperative binding* in vivo. http://biorxiv.org/lookup/doi/10.1101/2020.06.29.167155 (2020) doi:10.1101/2020.06.29.167155.
126. Lee, I. *et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing*. http://biorxiv.org/lookup/doi/10.1101/504993 (2018) doi:10.1101/504993.
127. Wang, Y. *et al.* Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* **29**, 1329–1342 (2019).
128. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
129. Abdulhay, N. J. *et al. Massively multiplex single-molecule oligonucleosome footprinting*. http://biorxiv.org/lookup/doi/10.1101/2020.05.20.105379 (2020) doi:10.1101/2020.05.20.105379.
130. Gargiulo, G. *et al.* NA-Seq: a discovery tool for the analysis of chromatin structure and dynamics during differentiation. *Dev. Cell* **16**, 466–481 (2009).
131. Chen, P. B., Zhu, L. J., Hainer, S. J., McCannell, K. N. & Fazzio, T. G. Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo. *BMC Genomics* **15**, 1104 (2014).
132. Chereji, R. V., Eriksson, P. R., Ocampo, J., Prajapati, H. K. & Clark, D. J. Accessibility of promoter DNA is not the primary determinant of chromatin-mediated gene regulation. *Genome Res.* **29**, 1985–1995 (2019).
133. Oberbeckmann, E. *et al.* Absolute nucleosome occupancy map for the Saccharomyces cerevisiae genome. *Genome Res.* **29**, 1996–2009 (2019).
134. Ponnaluri, V. K. C. *et al.* NicE-seq: high resolution open chromatin profiling. *Genome Biol.* **18**, 122 (2017).
135. Giresi, P. G. & Lieb, J. D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods San Diego Calif* **48**, 233–239 (2009).
136. Lai, B. *et al.* Trac-looping measures genome structure and chromatin accessibility. *Nat. Methods* **15**, 741–747 (2018).
137. Spracklin, G. & Pradhan, S. Protect-seq: genome-wide profiling of nuclease inaccessible domains reveals physical properties of chromatin. *Nucleic Acids Res.* **48**, e16 (2020).
138. Tchasovnikarova, I. A. *et al.* Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. *Nat. Genet.* **49**, 1035–1044 (2017).
139. Timms, R. T., Tchasovnikarova, I. A. & Lehner, P. J. Differential viral accessibility (DIVA) identifies alterations in chromatin architecture through large-scale mapping of lentiviral integration sites. *Nat. Protoc.* **14**, 153–170 (2019).
140. Aughey, G. N., Estacio Gomez, A., Thomson, J., Yin, H. & Southall, T. D. CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. *eLife* **7**, (2018).

141. Umeyama, T. & Ito, T. DMS-Seq for In Vivo Genome-wide Mapping of Protein-DNA Interactions and Nucleosome Centers. *Cell Rep.* **21**, 289–300 (2017).
142. Ishii, H., Kadonaga, J. T. & Ren, B. MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3457-3465 (2015).
143. Flaus, A., Luger, K., Tan, S. & Richmond, T. J. Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1370–1375 (1996).
144. Brogaard, K., Xi, L., Wang, J.-P. & Widom, J. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**, 496–501 (2012).
145. Voong, L. N. *et al.* Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell* **167**, 1555-1570.e15 (2016).
146. Chereji, R. V., Ramachandran, S., Bryson, T. D. & Henikoff, S. Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biol.* **19**, 19 (2018).
147. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
148. Chen, X., Miragaia, R. J., Natarajan, K. N. & Teichmann, S. A. A rapid and robust method for single cell chromatin accessibility profiling. *Nat. Commun.* **9**, 5345 (2018).
149. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
150. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **37**, 916–924 (2019).
151. 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).
152. Mezger, A. *et al.* High-throughput chromatin accessibility profiling at single-cell resolution. *Nat. Commun.* **9**, 3647 (2018).
153. Ma, S. *et al. Chromatin potential identified by shared single cell profiling of RNA and chromatin*. http://biorxiv.org/lookup/doi/10.1101/2020.06.17.156943 (2020) doi:10.1101/2020.06.17.156943.
154. Yin, Y. *et al.* High-Throughput Single-Cell Sequencing with Linear Amplification. *Mol. Cell* **76**, 676-690.e10 (2019).
155. Zhu, C. *et al.* An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat. Struct. Mol. Biol.* **26**, 1063–1070 (2019).
156. Lee, J. *et al. Kundajelab/Atac_Dnase_Pipelines: 0.3.0*. (Zenodo, 2016). doi:10.5281/ZENODO.156534.
157. Babraham Bioinformatics. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
158. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinforma. Oxf. Engl.* **32**, 3047–3048 (2016).
159. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
160. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).
161. Aronesty, E. *ea-utils : 'Command-line tools for processing biological sequencing data'*. (2011).
162. Cooper, J., Ding, Y., Song, J. & Zhao, K. Genome-wide mapping of DNase I hypersensitive sites in rare cell populations using single-cell DNase sequencing. *Nat. Protoc.* **12**, 2342–2354 (2017).
163. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
164. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
165. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
166. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).

167. Ou, J. *et al.* ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics* **19**, 169 (2018).
168. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
169. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
170. Buels, R. *et al.* JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).
171. Hofmeister, B. T. & Schmitz, R. J. Enhanced JBrowse plugins for epigenomics data visualization. *BMC Bioinformatics* **19**, 159 (2018).
172. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
173. Rashid, N. U., Giresi, P. G., Ibrahim, J. G., Sun, W. & Lieb, J. D. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. *Genome Biol.* **12**, R67 (2011).
174. Tarbell, E. D. & Liu, T. HMMRATAC: a Hidden Markov ModeleR for ATAC-seq. *Nucleic Acids Res.* **47**, e91–e91 (2019).
175. Boyle, A. P., Guinney, J., Crawford, G. E. & Furey, T. S. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinforma. Oxf. Engl.* **24**, 2537–2538 (2008).
176. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264–268 (2011).
177. Zhou, X., Blocker, A. W., Airoldi, E. M. & O'Shea, E. K. A computational approach to map nucleosome positions and alternative chromatin states with base pair resolution. *eLife* **5**, e16970 (2016).
178. Chen, K. *et al.* DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351 (2013).
179. Schep, A. N. *et al.* Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* **25**, 1757–1770 (2015).
180. Koohy, H., Down, T. A., Spivakov, M. & Hubbard, T. A Comparison of Peak Callers Used for DNase-Seq Data. *PLoS ONE* **9**, e96303 (2014).
181. Samb, R. *et al.* Using informative Multinomial-Dirichlet prior in a t-mixture with reversible jump estimation of nucleosome positions for genome-wide profiling. *Stat. Appl. Genet. Mol. Biol.* **14**, (2015).
182. The ENCODE Project Consortium *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
183. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
184. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
185. 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).
186. 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).
187. Stark, R. & Brown, G. *DiffBind: differential binding analysis of ChIP-Seq peak data.* (2011).
188. 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).
189. Liang, K. & Keles, S. Detecting differential binding of transcription factors with ChIP-seq. *Bioinforma. Oxf. Engl.* **28**, 121–122 (2012).
190. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016).
191. Gandolfi, F. & Tramontano, A. A computational approach for the functional classification of the epigenome. *Epigenetics Chromatin* **10**, 26 (2017).
192. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
193. 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).

194. Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
195. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
196. Herrmann, C., Van De Sande, B., Potier, D. & Aerts, S. i-cisTarget: An integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res.* **40**, (2012).
197. Imrichová, H., Hulselmans, G., Kalender Atak, Z., Potier, D. & Aerts, S. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res.* **43**, W57–W64 (2015).
198. Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinforma. Oxf. Engl.* **32**, 587–589 (2016).
199. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
200. Mammana, A. & Chung, H.-R. Chromatin segmentation based on a probabilistic model for read counts explains a large portion of the epigenome. *Genome Biol.* **16**, 151 (2015).
201. Hoffman, M. M. *et al.* Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* **9**, 473–476 (2012).
202. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
203. Stormo, G. D., Schneider, T. D., Gold, L. & Ehrenfeucht, A. Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli. *Nucleic Acids Res.* **10**, 2997–3011 (1982).
204. Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* gkz1001 (2019) doi:10.1093/nar/gkz1001.
205. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
206. Wingender, E., Dietze, P., Karas, H. & Knüppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res.* **24**, 238–241 (1996).
207. Kulakovskiy, I. V. *et al.* HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).
208. Thomas-Chollier, M. *et al.* RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res.* **40**, e31–e31 (2012).
209. Pavesi, G., Mereghetti, P., Mauri, G. & Pesole, G. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.* **32**, W199–W203 (2004).
210. Hiranuma, N., Lundberg, S. & Lee, S.-I. DeepATAC: A deep-learning method to predict regulatory factor binding activity from ATAC-seq signals. 1–5 (2017) doi:10.1101/172767.
211. Shrikumar, A., Greenside, P. & Kundaje, A. Learning Important Features Through Propagating Activation Differences. *ArXiv170402685 Cs* (2017).
212. Minnoye, L. *et al. Cross-species analysis of melanoma enhancer logic using deep learning*. http://biorxiv.org/lookup/doi/10.1101/2019.12.21.885715 (2019) doi:10.1101/2019.12.21.885715.
213. Mariani, L., Weinand, K., Gisselbrecht, S. S. & Bulyk, M. L. MEDEA: analysis of transcription factor binding motifs in accessible chromatin. *Genome Res.* **30**, 736–748 (2020).
214. Baek, S. & Sung, M.-H. Genome-Scale Analysis of Cell-Specific Regulatory Codes Using Nuclear Enzymes. in *Statistical Genomics* (eds. Mathé, E. & Davis, S.) vol. 1418 225–240 (Springer New York, 2016).
215. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
216. Schwessinger, R. *et al.* Sasquatch: predicting the impact of regulatory SNPs on transcription factor binding from cell- and tissue-specific DNase footprints. *Genome Res.* **27**, 1730–1742 (2017).
217. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).

218. Piper, J. *et al.* Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res.* **41**, e201 (2013).
219. Gusmao, E. G., Dieterich, C., Zenke, M. & Costa, I. G. Detection of active transcription factor binding sites with the combination of DNase hypersensitivity and histone modifications. *Bioinformatics* **30**, 3143–3151 (2014).
220. Chen, X., Hoffman, M. M., Bilmes, J. A., Hesselberth, J. R. & Noble, W. S. A dynamic Bayesian network for identifying protein-binding footprints from single molecule-based sequencing data. *Bioinformatics* **26**, i334–i342 (2010).
221. Sung, M.-H., Guertin, M. J., Baek, S. & Hager, G. L. DNase footprint signatures are dictated by factor dynamics and DNA sequence. *Mol. Cell* **56**, 275–285 (2014).
222. Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21**, 447–455 (2011).
223. Yardımcı, G. G., Frank, C. L., Crawford, G. E. & Ohler, U. Explicit DNase sequence bias modeling enables high-resolution transcription factor footprint detection. *Nucleic Acids Res.* **42**, 11865–11878 (2014).
224. Vierstra, J. & Stamatoyannopoulos, J. A. Genomic footprinting. *Nat. Methods* **13**, 213–221 (2016).
225. Quach, B. & Furey, T. S. DeFCoM: analysis and modeling of transcription factor binding sites using a motif-centric genomic footprinter. *Bioinforma. Oxf. Engl.* **33**, 956–963 (2017).
226. Sung, M.-H., Baek, S. & Hager, G. L. Genome-wide footprinting: ready for prime time? *Nat. Methods* **13**, 222–228 (2016).
227. Chen, H. *et al.* Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biol.* **20**, 241 (2019).
228. Baek, S. & Lee, I. Single-cell ATAC sequencing analysis: From data preprocessing to hypothesis generation. *Comput. Struct. Biotechnol. J.* **18**, 1429–1439 (2020).
229. Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* btz625 (2019) doi:10.1093/bioinformatics/btz625.
230. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
231. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
232. Luecken, M. *et al. Benchmarking atlas-level data integration in single-cell genomics*. http://biorxiv.org/lookup/doi/10.1101/2020.05.22.111161 (2020) doi:10.1101/2020.05.22.111161.
233. Jin, S., Zhang, L. & Nie, Q. scAI: an unsupervised approach for the integrative analysis of parallel single-cell transcriptomic and epigenomic profiles. *Genome Biol.* **21**, 25 (2020).
234. Chen, H. *et al.* Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat. Commun.* **10**, 1903 (2019).
235. Trapnell, C. & Cacchiarelli, D. *Monocle : Cell counting , differential expression , and trajectory analysis for single-cell RNA-Seq experiments*. (2016).
236. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
237. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
238. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
239. Gaffney, D. J. *et al.* Controls of nucleosome positioning in the human genome. *PLoS Genet.* **8**, e1003036 (2012).
240. Tillo, D. *et al.* High nucleosome occupancy is encoded at human regulatory sequences. *PloS One* **5**, e9129 (2010).
241. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7**, e1002266 (2011).
242. Schulz, K. N. *et al.* Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res.* **25**, 1715–1726 (2015).

243. Sun, Y. *et al.* Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. *Genome Res.* **25**, 1703–1714 (2015).
244. Gao, L. *et al.* Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell* 1–12 (2018) doi:10.1016/j.cell.2018.02.028.
245. Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).
246. Leichsenring, M., Maes, J., Mössner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005–1009 (2013).
247. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **293**, 13795–13804 (2018).
248. Fernandez Garcia, M. *et al.* Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol. Cell* **75**, 921-932.e6 (2019).
249. Iwafuchi, M. *et al.* Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat. Genet.* **52**, 418–427 (2020).
250. Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22534–22539 (2010).
251. Svaren, J., Klebanow, E., Sealy, L. & Chalkley, R. Analysis of the competition between nucleosome formation and transcription factor binding. *J. Biol. Chem.* **269**, 9335–9344 (1994).
252. Workman, J. L. & Kingston, R. E. Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. *Science* **258**, 1780–1784 (1992).
253. Lone, I. N. *et al.* Binding of NF-κB to nucleosomes: effect of translational positioning, nucleosome remodeling and linker histone H1. *PLoS Genet.* **9**, e1003830 (2013).
254. Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4**, e1000138 (2008).
255. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011).
256. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
257. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463**, 474–484 (2010).
258. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18 (2011).
259. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).
260. Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548.e16 (2018).
261. Yoshida, H. *et al.* The cis-Regulatory Atlas of the Mouse Immune System. *Cell* **176**, 897-912.e20 (2019).
262. Gosselin, D. *et al.* Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
263. Lavin, Y. *et al.* Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
264. Satpathy, A. T. *et al.* Transcript-indexed ATAC-seq for precision immune profiling. *Nat. Med.* **24**, 580–590 (2018).
265. Wei, G. *et al.* Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. *Immunity* **30**, 155–167 (2009).
266. Koues, O. I. *et al.* Distinct Gene Regulatory Pathways for Human Innate versus Adaptive Lymphoid Cells. *Cell* **165**, 1134–1146 (2016).
267. Shih, H.-Y. *et al.* Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality. *Cell* **165**, 1120–1133 (2016).
268. Youngblood, B. *et al.* Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature* **552**, 404–409 (2017).
269. van der Veeken, J. *et al.* Memory of Inflammation in Regulatory T Cells. *Cell* **166**, 977–990 (2016).

270. Netea, M. G. *et al.* Defining trained immunity and its role in health and disease. *Nat. Rev. Immunol.* **20**, 375–388 (2020).
271. Novakovic, B. *et al.* β-Glucan Reverses the Epigenetic State of LPS-Induced Immunological Tolerance. *Cell* **167**, 1354-1368.e14 (2016).
272. Saeed, S. *et al.* Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**, 1251086 (2014).
273. Krausgruber, T. *et al.* Structural cells are key regulators of organ-specific immune responses. *Nature* **583**, 296–302 (2020).
274. de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289-304.e18 (2018).
275. Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
276. Trevino, A. E. *et al.* Chromatin accessibility dynamics in a model of human forebrain development. *Science* **367**, (2020).
277. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
278. Yin, S. *et al.* Transcriptomic and open chromatin atlas of high-resolution anatomical regions in the rhesus macaque brain. *Nat. Commun.* **11**, 474 (2020).
279. Jia, G. *et al.* Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nat. Commun.* **9**, 4877 (2018).
280. Stone, N. R. *et al.* Context-Specific Transcription Factor Functions Regulate Epigenomic and Transcriptional Dynamics during Cardiac Reprogramming. *Cell Stem Cell* **25**, 87-102.e9 (2019).
281. Fan, X. *et al.* Single Cell and Open Chromatin Analysis Reveals Molecular Origin of Epidermal Cells of the Skin. *Dev. Cell* **47**, 21-37.e5 (2018).
282. Dravis, C. *et al.* Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. *Cancer Cell* **34**, 466-482.e6 (2018).
283. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).
284. Beekman, R. *et al.* The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.* **24**, 868–880 (2018).
285. Ott, C. J. *et al.* Enhancer Architecture and Essential Core Regulatory Circuitry of Chronic Lymphocytic Leukemia. *Cancer Cell* **34**, 982-995.e7 (2018).
286. Rendeiro, A. F. *et al.* Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nat. Commun.* **11**, 577 (2020).
287. Yi, G. *et al.* Chromatin-Based Classification of Genetically Heterogeneous AMLs into Two Distinct Subtypes with Diverse Stemness Phenotypes. *Cell Rep.* **26**, 1059-1069.e6 (2019).
288. Schmidl, C. *et al.* Combined chemosensitivity and chromatin profiling prioritizes drug combinations in CLL. *Nat. Chem. Biol.* **15**, 232–240 (2019).
289. Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).
290. Akhtar-Zaidi, B. *et al.* Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**, 736–739 (2012).
291. Cohen, A. J. *et al.* Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat. Commun.* **8**, 14400 (2017).
292. Guilhamon, P. *et al. Single-cell chromatin accessibility in glioblastoma delineates cancer stem cell heterogeneity predictive of survival*. http://biorxiv.org/lookup/doi/10.1101/370726 (2018) doi:10.1101/370726.
293. Tome-Garcia, J. *et al.* Analysis of chromatin accessibility uncovers TEAD1 as a regulator of migration in human glioblastoma. *Nat. Commun.* **9**, 4020 (2018).
294. Ooi, W. F. *et al.* Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat. Commun.* **7**, 12983 (2016).
295. Denny, S. K. *et al.* Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility. *Cell* **166**, 328–342 (2016).
296. Wang, Z. *et al.* The Open Chromatin Landscape of Non-Small Cell Lung Carcinoma. *Cancer Res.* **79**, 4840–4854 (2019).

297. Riggi, N. *et al.* EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* **26**, 668–681 (2014).
298. Tomazou, E. M. *et al.* Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. *Cell Rep.* **10**, 1082–1095 (2015).
299. Torchia, J. *et al.* Integrated (epi)-Genomic Analyses Identify Subgroup-Specific Therapeutic Targets in CNS Rhabdoid Tumors. *Cancer Cell* **30**, 891–908 (2016).
300. Halbritter, F. *et al.* Epigenomics and Single-Cell Sequencing Define a Developmental Hierarchy in Langerhans Cell Histiocytosis. *Cancer Discov.* **9**, 1406–1421 (2019).
301. Miller, B. C. *et al.* Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
302. Sen, D. R. *et al.* The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016).
303. Ghoneim, H. E. *et al.* De Novo Epigenetic Programs Inhibit PD-1 Blockade-Mediated T Cell Rejuvenation. *Cell* **170**, 142-157.e19 (2017).
304. Pauken, K. E. *et al.* Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).
305. Waldman, A. D., Fritz, J. M. & Lenardo, M. J. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat. Rev. Immunol.* (2020) doi:10.1038/s41577-020-0306-5.
306. Peng, D. *et al.* Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. *Nature* **527**, 249–253 (2015).
307. Boyd, M. *et al.* Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies. *Nat. Commun.* **9**, 1661 (2018).
308. Ai, R. *et al.* Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes. *Nat. Commun.* **9**, 1921 (2018).
309. Klein, H.-U. *et al.* Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat. Neurosci.* **22**, 37–46 (2019).
310. Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, 3121 (2018).
311. Sun, W. *et al.* Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* **167**, 1385-1397.e11 (2016).
312. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22**, 1748–1759 (2012).
313. Xiao, Y., Liu, H., Wu, L., Warburton, M. & Yan, J. Genome-wide Association Studies in Maize: Praise and Stargaze. *Mol. Plant* **10**, 359–374 (2017).
314. Degner, J. F. *et al.* DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012).
315. Gate, R. E. *et al.* Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* **50**, 1140–1150 (2018).
316. Jacobs, J. *et al.* The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat. Genet.* **50**, 1011–1020 (2018).
317. Atak, Z. K. *et al. Prioritization of enhancer mutations by combining allele-specific chromatin accessibility with deep learning*. http://biorxiv.org/lookup/doi/10.1101/2019.12.21.885806 (2019) doi:10.1101/2019.12.21.885806.
318. Roscito, J. G. *et al.* Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. *Nat. Commun.* **9**, 4737 (2018).
319. Stone, J. R. & Wray, G. A. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.* **18**, 1764–1770 (2001).
320. Van de Velde, J., Van Bel, M., Vaneechoutte, D. & Vandepoele, K. A Collection of Conserved Noncoding Sequences to Study Gene Regulation in Flowering Plants. *Plant Physiol.* **171**, 2586–2598 (2016).
321. Lu, Z. *et al.* The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* **5**, 1250–1259 (2019).

322. Maher, K. A. *et al.* Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. *Plant Cell* **30**, 15–36 (2018).
323. Sebé-Pedrós, A. *et al.* The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. *Cell* **165**, 1224–1237 (2016).
324. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* **41**, D991-995 (2013).
325. Kolesnikov, N. *et al.* ArrayExpress update--simplifying data submissions. *Nucleic Acids Res.* **43**, D1113-1116 (2015).
326. Leinonen, R., Sugawara, H., Shumway, M. & International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* **39**, D19-21 (2011).
327. Leinonen, R. *et al.* The European Nucleotide Archive. *Nucleic Acids Res.* **39**, D28-31 (2011).
328. Kaminuma, E. *et al.* DDBJ launches a new archive database with analytical tools for next-generation sequence data. *Nucleic Acids Res.* **38**, D33-38 (2010).
329. Lappalainen, I. *et al.* The European Genome-phenome Archive of human data consented for biomedical research. *Nat. Genet.* **47**, 692–695 (2015).
330. Mailman, M. D. *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39**, 1181–1186 (2007).
331. Davie, K. *et al.* A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174**, 982-998.e20 (2018).
332. David, F. P. A., Litovchenko, M., Deplancke, B. & Gardeux, V. ASAP 2020 update: an open, scalable and interactive web-based portal for (single-cell) omics analyses. *Nucleic Acids Res.* **48**, W403–W414 (2020).
333. Avsec, Ž. *et al.* The Kipoi repository accelerates community exchange and reuse of predictive models for genomics. *Nat. Biotechnol.* **37**, 592–600 (2019).
334. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
335. Hah, N., Murakami, S., Nagari, A., Danko, C. G. & Kraus, W. L. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* **23**, 1210–1223 (2013).
336. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394 (2011).
337. Berest, I. *et al.* Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. *Cell Rep.* **29**, 3147-3159.e12 (2019).
338. Colli, M. L. *et al.* An integrated multi-omics approach identifies the landscape of interferon-α-mediated responses of human pancreatic beta cells. *Nat. Commun.* **11**, 2584 (2020).
339. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).
340. Chen, X. *et al.* Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation of epigenomic heterogeneity. *Nat. Commun.* **9**, 4590 (2018).
341. Pott, S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* **6**, (2017).
342. Guo, F. *et al.* Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res.* **27**, 967–988 (2017).
343. Lhoumaud, P. *et al.* EpiMethylTag: simultaneous detection of ATAC-seq or ChIP-seq signals with DNA methylation. *Genome Biol.* **20**, 248 (2019).
344. Spektor, R., Tippens, N. D., Mimoso, C. A. & Soloway, P. D. methyl-ATAC-seq measures DNA methylation at accessible chromatin. *Genome Res.* **29**, 969–977 (2019).
345. Barnett, K. R. *et al.* ATAC-Me Captures Prolonged DNA Methylation of Dynamic Chromatin Accessibility Loci during Cell Fate Transitions. *Mol. Cell* **77**, 1350-1364.e6 (2020).
346. Clark, S. J. *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).
347. Argelaguet, R. *et al.* Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, (2018).

348. Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176**, 361-376.e17 (2019).
349. Chen, X. *et al.* ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
350. Thornton, C. A. *et al. Spatially-mapped single-cell chromatin accessibility*. http://biorxiv.org/lookup/doi/10.1101/815720 (2019) doi:10.1101/815720.