Check for updates

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]✉

Abstract | Chromatin accessibility, or the physical access to chromatinized DNA, is a widely studied characteristic of the eukaryotic genome. As active regulatory DNA elements are generally 'accessible', the genome-wide profiling of chromatin accessibility can be used to identify candidate regulatory genomic regions in a tissue or cell type. Multiple biochemical methods have been developed to profile chromatin accessibility, both in bulk and at the single-cell level. Depending on the method, enzymatic cleavage, transposition or DNA methyltransferases are used, followed by high-throughput sequencing, providing a view of genome-wide chromatin accessibility. In this Primer, we discuss these biochemical methods, as well as bioinformatics tools for analysing and interpreting the generated data, and insights into the key regulators underlying developmental, evolutionary and disease processes. We outline standards for data quality, reproducibility and deposition used by the genomics community. Although chromatin accessibility profiling is invaluable to study gene regulation, alone it provides only a partial view of this complex process. Orthogonal assays facilitate the interpretation of accessible regions with respect to enhancer–promoter proximity, functional transcription factor binding and regulatory function. We envision that technological improvements including single-molecule, multi-omics and spatial methods will bring further insight into the secrets of genome regulation.

Nucleosomes
The basic structural unit of DNA packaging, consisting of ~147 bp of DNA wrapped around an octamer of histones.

Cis-regulatory elements
Non-coding DNA regions involved in the regulation of expression of neighbouring genes. The regions contain binding sites for transcription factors.

Accessible chromatin
A permissive state of the chromatin in which nuclear macromolecules are able to physically access and interact with the DNA.

✉*e-mail: stein.aerts@kuleuven.vib.be*

https://doi.org/10.1038/s43586-020-00008-9

Chromatin accessibility refers to the level of physical compaction of chromatin, a complex formed by DNA and associated proteins consisting mainly of histones, transcription factors (TFs), chromatin-modifying enzymes and chromatin-remodelling complexes[1–3]. Although eukaryotic genomes are generally packed into nucleosomes, which comprise ~147 bp of DNA wrapped around an octamer of histones[4,5], nucleosome occupancy is not uniform in the genome, and varies across tissues and cell types. Nucleosomes are typically depleted at genomic locations that represent *cis*-regulatory elements — enhancers and promoters, among others — that interact with transcriptional regulators (for example, TFs), resulting in accessible chromatin[6–10]. Profiling chromatin accessibility on a genome-wide scale is an excellent tool to map putative regulatory elements in a cell type or cell state.

Post-translational chemical modifications of chromatin, including DNA methylation (in vertebrates) and histone methylation and acetylation, are dynamic and change between different cell states, similar to nucleosome positioning. These post-translational modifications are often correlated with chromatin accessibility and can reflect specific functionalities of genomic regions related to the regulation of gene expression[11,12]. Changes in these post-translational modifications, such as increased or decreased histone methylation and acetylation, are affected by a large set of chromatin-modifying enzymes that can be recruited to chromatin regions by TFs. These modifications alter the physico-chemical properties of the chromatin, which in turn can influence the formation of transcriptional condensates[13,14]. In addition, active chromatin remodelling impacts nucleosome occupancy; for example, the SWI/SNF complexes use ATP hydrolysis to alter histone–DNA contacts, thereby repositioning or removing nucleosomes[15]. Dynamic changes in the chromatin structure, chemical modifications and nucleosome positioning form a crucial interplay with the TFs that drive differentiation of cells during development[16,17]. Initial changes in chromatin accessibility are caused by the binding of TFs, which outcompete histones and recruit cofactors, including ATP-dependent chromatin remodellers[18,19], or by TFs that preferentially bind to their recognition sequence in nucleosomal DNA[20,21]. The binding of these initial TFs, known as pioneer factors, can recruit other TFs to co-bind and further stabilize the nucleosome-depleted

Author addresses

[1]Center for Brain & Disease Research, VIB-KU Leuven, Leuven, Belgium.
[2]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, Bethesda, MD, USA.
[6]Department of Genetics, University of Georgia, Athens, GA, USA.
[7]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany.
[8]Institute of Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria.

region, further contributing to the regulation of gene expression of target genes[22–24]. Consequently, the analysis of TF binding sites in regulatory regions within accessible chromatin can bring insights into cell type-specific lineage factors and gene regulatory networks.

Various changes in the chromatin landscape, as well as mutations in chromatin remodellers and in regulatory regions, are linked to a range of traits and diseases[25–28]. In fact, many causal genome-wide association study variants are located in accessible regulatory elements[29]. In order to improve our understanding of chromatin dynamics during development and in disease contexts, researchers and large consortia, including the Encyclopedia of DNA Elements (ENCODE) Consortium[30], the International Human Epigenome Consortium (IHEC)[31], the National Institutes of Health (NIH) Roadmap Epigenomics Mapping Consortium[32] and the BLUEPRINT epigenome project[33], have collected and compared chromatin landscapes across cell types and during disease development.

Over the past decades, several chromatin accessibility profiling methods have been developed and widely used[34–44]. Generally, these methods are based on the physical accessibility of the chromatin to enzymes, which mark the accessible DNA by fragmentation, tagmentation or chemical labelling (for example, methylation of GpC dinucleotides). Initial research in the 1970s showed that regions of active transcription, such as promoters and introns of expressed genes, are particularly sensitive to digestion by DNA endonucleases such as deoxyribonuclease I (DNase I), indicative of a particularly accessible form of the chromatin[45]. Moreover, chromatin is digested at regularly spaced sites due to nucleosome positioning[2,46]. DNase I is still 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[47–49]. With the advent of next-generation sequencing (NGS) techniques, DNase I hypersensitive site sequencing (DNase-seq) was one of the first adaptations to perform genome-wide profiling of accessible chromatin[35,40], which was followed by a handful of other methods. Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and variants[36–38] together with DNase-seq are the two most commonly used chromatin accessibility profiling methods today[50].

Importantly, as regulatory regions co-define a cell type, their chromatin accessibility is cell type-dependent[10,51–53]. When investigating heterogeneous samples, it is therefore advisable to measure chromatin accessibility in isolated subpopulations — by flow cytometry-based purification[53] — or at the single-cell level to avoid averaging over heterogeneous cell populations (FIG. 1). Currently, the field of single-cell omics, including single-cell epigenomic assays such as single-cell ATAC-seq (scATAC-seq) and single-cell DNase-seq (scDNase-seq), provides exciting new opportunities to study genome regulation in complex tissues such as the brain, whole embryos and tumours[54–61]. Accompanied by the rise of single-cell chromatin accessibility profiling, a wide range of bioinformatics tools have been developed that allow analysis of the generated data, which is intrinsically sparse due to experimental limitations. Indeed, single-cell epigenomics is technically more challenging compared with single-cell transcriptomics because most loci in a diploid cell are present in only two copies of DNA that can be assayed.

Although chromatin accessibility profiling methods may serve as an analytic foundation to identify regulatory regions, it is reported that often less than 50% of accessible regions in human DNA are active as enhancers[62,63]. Interestingly, however, work in both the *Drosophila* embryo[55] and the *Drosophila* eye imaginal disc[64] shows that when a genomic region is uniquely accessible in a specific cell type, more than 80% of the accessible enhancers are also active in the corresponding cell type[55,64,65]. In addition, linking active accessible regulatory regions to their target genes solely based on accessibility data remains a challenge. Therefore, additional data, including those from transcriptomics, enhancer–reporter assays and 3D chromatin conformation maps, help to determine the function of an accessible region and identify its putative target genes, especially when combined in a multi-omics way[64,66–71].

This Primer provides an overview of the commonly used and most recently developed chromatin accessibility profiling methods, both in bulk and at the single-cell level (Experimentation). In addition, it provides an outline of computational analysis techniques (Results) and examples of use in diverse organisms and fields (Applications). Finally, the Primer discusses standards for data sharing (Reproducibility and data deposition), and examines currently unmet needs (Limitations and optimizations) and future opportunities for technological development (Outlook) that will increase our understanding of chromatin accessibility landscapes and their functional role in gene regulation during development, during evolution and in disease contexts.

Experimentation

Bulk 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[46]. However, NGS has revolutionized the way chromatin is investigated by allowing us to study its accessibility genome-wide. In this section, we will briefly describe the principles and the pros and cons of several commonly used experimental techniques to assess chromatin accessibility or nucleosome positioning in bulk, including DNase-seq, ATAC-seq and micrococcal

Transcriptional condensates
Membraneless compartments of the genome formed by liquid–liquid phase separation, in which the transcription machinery is concentrated to efficiently activate transcription.

Pioneer factors
Transcription factors that can recognize and bind their target sequence in closed chromatin and trigger opening of the chromatin, allowing binding of other transcription factors.

Tagmentation
Transposases cut DNA into fragments while simultaneously adding adaptor sequences. Used in Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), as well as for general sequencing library construction to randomly fragment double-stranded DNA.

TF footprinting
Small stretches of nucleotides that are protected from cleavage or tagmentation and represent the location of transcription factor (TF) binding sites. TF footprints can be inferred from the analysis of high-resolution chromatin accessibility data.

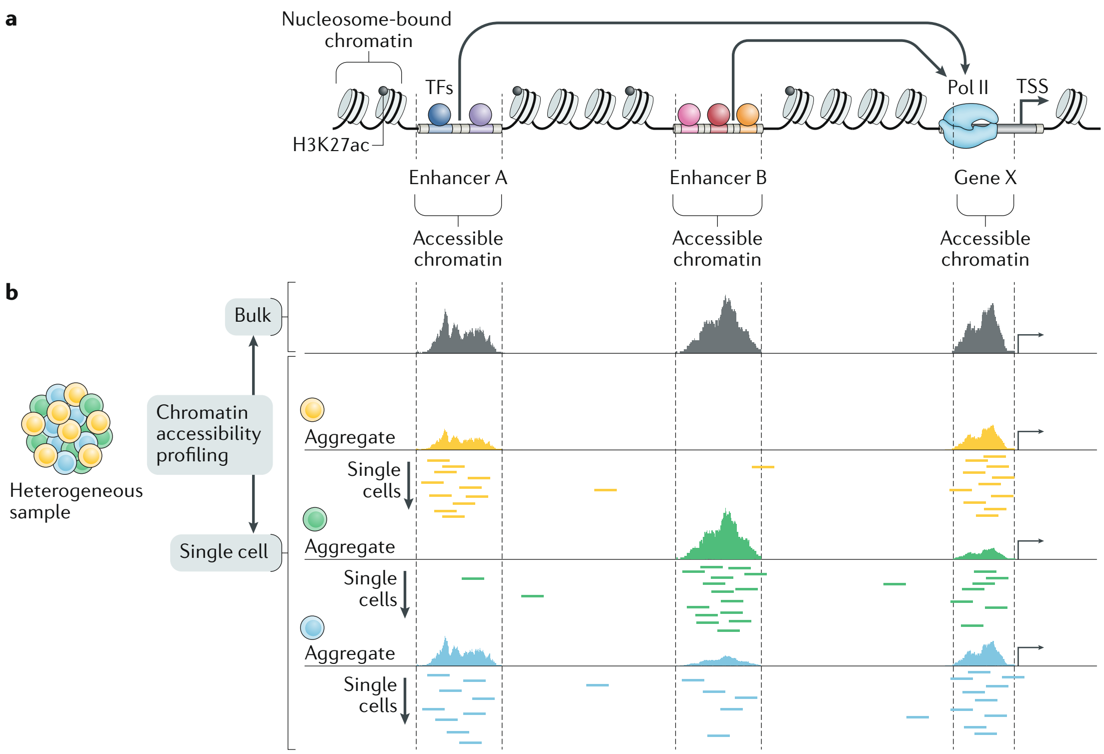

Fig. 1 | **Chromatin accessibility profiling in bulk and at single-cell level reveals putative regulatory regions.** **a** | Representation of a chromatin landscape is shown in which transcription factor (TF)-bound enhancers and the promoter of a gene are nucleosome depleted and thus accessible. The TFs are represented as coloured circles and the arrows represent 3D looping of the enhancers to the promoter of the target gene. **b** | Bulk and single-cell chromatin accessibility profiles of a heterogeneous sample containing three different cell populations. When performing single-cell chromatin accessibility profiling, sparse single-cell data are used to cluster cells, often followed by aggregating the reads per cluster, thereby reconstituting pseudo-bulk profiles per cluster or cell type. H3K27ac, histone H3 acetylated at lysine 27; Pol II, polymerase II; TSS, transcription start site.

nuclease sequencing (MNase-seq) as well as several single-molecule chromatin accessibility profiling methods (TABLE 1). In addition, we discuss chromatin immunoprecipitation followed by sequencing (ChIP–seq) and related methods, as these are powerful techniques to gain further insights into chromatin landscapes and TF binding. Finally, various less commonly used chromatin accessibility and nucleosome positioning methods are described in BOX 1.

DNase-seq. One of the first genome-wide profiling experiments of accessible chromatin was published 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[35,40] (FIG. 2a). In DNase-seq, nuclei are first isolated and permeabilized using a mild detergent such as 0.1% Triton X-100, such that the DNase I enzyme can enter the nucleus efficiently. After digestion, the small DNA fragments (50–100 bp) are purified and size-selected for downstream library construction and sequencing. As DNase I digestion is an enzymatic process, the amount of the enzyme can significantly affect the digestion efficiency and, thereby, also the quality of the data. Therefore, 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. In addition to fresh cells, DNase-seq has also been performed on formalin-fixed paraffin-embedded samples[10,35].

Beyond this requirement of careful enzyme titrations, major limitations of the traditional DNase-seq assay include the large number of cells (tens of millions) required as input material and its tedious and lengthy protocol that takes several days[72]. However, recently, a modified DNase-seq assay (scDNase-seq) has been developed to analyse single cells or a small number of cells[58,73]. scDNase-seq requires only hundreds to thousands of either fresh or fixed cells for a bulk cell assay and takes 1 day for library construction, without the need for fractionation of DNA fragments[74]. Caution must be taken when interpreting DNase-seq results because they show some intrinsic bias in cleavage sites. The DNA minor groove shows variation in width depending on the sequence, and a narrower groove is preferentially cleaved by DNase I. Also, CpG methylation enhances adjacent DNase I cleavage[75,76]. These factors should be considered when interpreting the footprint of a TF[77]. DNase-seq is the method of choice to detect TF footprints[78].

ATAC-seq. ATAC-seq emerged as an alternative assay to investigate accessible chromatin profiles[36]. In this assay, a genetically engineered hyperactive DNA

Table 1 | **Comparison of the three most commonly used chromatin accessibility profiling methods**

Feature	DNase-seq	ATAC-seq	MNase-seq
Type of data produced	Accessible chromatin	Accessible chromatin	Nucleosomes/ inaccessible chromatin
Type of input	Fresh, fixed paraffin-embedded samples or formaldehyde cross-linked samples	Fresh, slowly cooled cryopreserved, snap-frozen or formaldehyde cross-linked samples	Fresh or formaldehyde cross-linked samples
Number of input cells[a]	1–10 million	500–50,000	10,000–100,000
Sequencing depth (for human samples)	20–50 million uniquely mapping reads (200 million for transcription factor footprinting analysis)	25 million non-mitochondrial uniquely mapping reads for standard analysis	150–200 million reads
Enzyme-specific cleavage bias?	Yes	Yes	Yes
Difficulty	Requires careful enzyme calibration and complicated sample preparations	Simple protocol, requires limited or no experimental calibration	Requires careful enzyme calibration
Time	Lengthy protocol (1–3 days)	Fast protocol (<1 day)	Lengthy protocol (2 days)

ATAC-seq, Assay for Transposase-Accessible Chromatin using sequencing; DNase-seq, deoxyribonuclease I (DNase I) hypersensitive site sequencing; MNase-seq, micrococcal nuclease sequencing. [a]Note that these numbers refer to the amount of input cells needed for standard bulk methods, but it is possible to go to a lower number of cells when using low-input or single-cell methods.

transposase (Tn5) preloaded with monovalent mosaic end adapters simultaneously cleaves and tags accessible or nucleosome-depleted chromatin regions[36,79,80] (FIG. 2b). The hyperactive Tn5 mutant includes three mutations that increase the relatively low activity of the wild-type Tn5 and allows for efficient in vitro integration of the mosaic end adapters[80]. The target DNA fragments are purified, PCR-amplified and sequenced by NGS. As both Tn5 transposase and DNase I recognize accessible chromatin, the sequences detected by ATAC-seq have been found to be highly enriched in DNase I hypersensitive sites (DHSs)[81–83].

A major advantage of ATAC-seq and its variants[37,38] is that they are very sensitive assays that work well on low-input samples (for example, 500–50,000 cells, compared with millions of cells for DNase-seq) and use a simpler protocol due to the simultaneous chromatin fragmentation and tagging[36]. If fresh cells are difficult to obtain, slowly cooled cryopreserved cells can also be profiled by ATAC-seq. In addition, it is possible to generate high signal to background profiles from formaldehyde-fixed cells using an adapted method[84], as well as from clinically relevant snap-frozen samples using the improved Omni-ATAC protocol[38] or from nuclei collected via flow cytometry[85]. The Omni-ATAC protocol improves the signal to noise ratio by mainly using a combination of multiple mild detergents to improve permeabilization across a wide array of cell types and to remove mitochondria from the reaction.

Similar to the enzyme-specific cleavage bias of DNase I, the Tn5 enzyme shows steric hindrance and sequence bias in chromatin tagmentation[79,86,87]. Accurate prediction of TF footprints from ATAC-seq data requires Tn5 bias correction that is different from DNase-seq bias correction[88]. An initial limitation of the first ATAC-seq protocol was the profiling of contaminating organellar DNA, such as mitochondrial DNA and/or chloroplast DNA for plants, or *Wolbachia* DNA in infected *Drosophila* stocks[36,89]. Large amounts of sequencing reads can be consumed by these contaminations, meaning that deeper sequencing is necessary to reach a good signal-to-noise ratio at regions of interest in the data. However, this limitation can be significantly reduced either by improved lysis conditions (as is the case in Omni-ATAC[38]), by purification of nuclei via flow cytometry[85] or by applying clustered regularly interspaced short palindromic repeats (CRISPR) technology to cleave mitochondrial ribosomal DNA prior to the experiment[82,90]. Another deficiency of the original procedure is that half of all fragments are lost as 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 using a T7 RNA polymerase linear amplification protocol[91].

Given the speed (a few hours) and straightforward nature of the protocol, combined with its sensitivity and requirement for low numbers of cells, ATAC-seq and its newer variants (for example, OmniATAC-seq) are currently the most commonly used methods to generate comprehensive chromatin accessibility maps in research laboratories (~400 data sets in PubMed in 2019 compared with <100 data sets for DNase-seq, MNase-seq and FAIRE-seq (formaldehyde-assisted isolation of regulatory elements) combined)[50]. In addition to profiling accessible chromatin, ATAC-seq can also be used to detect TF footprints and to map nucleosome positioning.

MNase-seq. Nucleosome positioning 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 (MNase-seq)[42] (FIG. 2c). MNase-seq is thus an orthogonal assay compared with DNase-seq and ATAC-seq as it measures nucleosome-occupied regions and is the most widely used method to map nucleosome positions genome-wide. A recently developed quantitative protocol for MNase-seq involves subjecting aliquots

Mosaic end adapters
Hyperactive versions of the two inverted 19-bp end sequences of the wild-type Tn5 transposon that, during an Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) experiment, are end-joined to accessible DNA by the transposase.

Nucleosome ladder
A characteristic 'ladder' pattern that originates from the cleavage of the linker DNA between nucleosomes, due to the periodic arrangement of nucleosomes.

of a sample to different levels of digestion by MNase, allowing clearer distinction of nucleosome positions and occupancy from higher-order chromatin properties, which can also be summarized in a theoretical framework[92].

In MNase-seq, 10,000–100,000 either fresh or formaldehyde cross-linked cells can be used for library construction. Digestion of chromatin by MNase typically results in a nucleosome ladder consisting of mononucleosome, dinucleosome, trinucleosome and so on, 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[42]. Similar to DNase-seq, MNase-seq requires careful enzyme titrations and is time-consuming (2-day protocol). Another limitation of MNase-seq is that it requires a large number of sequencing reads, preferably 150–200 million reads for human samples[93]. MNase-seq also suffers from enzyme-specific cleavage biases, specifically a preferential cleavage of A/T versus G/C[94].

MNase-seq has been applied to investigate the dynamics of the nucleosome landscape and their function in transcriptional regulation[95]. However, as nucleosome positioning 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[96].

Box 1 | **Less commonly used bulk chromatin accessibility profiling methods**

- Nicking enzyme assisted sequencing (NicE-seq)[370] uses a nicking enzyme to probe accessible DNA.
- Formaldehyde-assisted isolation of regulatory elements (FAIRE-seq)[39,371] profiles accessible chromatin based on its preferential release during sonication of cross-linked cells.
- Transposase-mediated analysis of chromatin looping (TrAC-looping)[372] uses Tn5 transposase and a bivalent mosaic end adaptor to detect genome-wide chromatin accessibility in addition to providing genome-wide chromatin interaction information on regulatory regions.
- Protect-seq[373] measures strongly heterochromatinized genomic regions, based on their resistance to nuclease digestion.
- Differential viral accessibility (DIVA)[374,375] uses preferential viral insertion into accessible DNA to map accessible chromatin regions.
- Chromatin accessibility profiling using targeted DamID (CATaDa)[376] labels open chromatin using ectopic expression of the *Escherichia coli* Dam methyltransferase.
- Dimethyl sulfate sequencing (DMS-seq)[377] probes protein-bound regions based on their escape from DMS attacks.
- Methidiumpropyl-EDTA sequencing (MPE-seq)[378] uses the chemical MPE-Fe(II) to map nucleosome positions.
- Nuclease-accessible site sequencing (NA-seq)[379] uses HpaII and NlaIII restriction enzymes to cleave and select for accessible sites in isolated nuclei.
- Restriction endonuclease digestion of chromatin coupled to deep sequencing (RED-seq)[380] is a modified NA-seq method, applicable to permeabilized cells.
- Quantitative DNA accessibility assay (qDA-seq)[381] uses restriction enzyme AluI to measure absolute accessibility and the rate at which accessible sites are cut.
- Occupancy measurement via restriction enzymes and high-throughput sequencing (ORE-seq)[382] uses restriction enzymes and has been applied to profile chromatin accessibility in yeast.
- Methods developed by Brogaard et al.[383], Voong et al.[384] and Chereji et al.[385] use chemical cleavage or modification reactions for direct mapping of nucleosome positions and are conceptually based on the original method by Flaus et al.[386].

Assays for single-molecule chromatin accessibility profiling. An emerging class of methods aim to map chromatin accessibility and TF binding in 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 fibres. The assays in this class rely on methyltransferase enzymes that preferentially modify accessible DNA (FIG. 2d). For years, the only read-out that such methods could rely on was bisulfite conversion of unmethylated cytosines followed by Sanger sequencing (for localized analysis of particular loci)[97–100] and, later, NGS (for both local and genome-wide coverage). The first genome-wide assay of this kind was the methylation accessibility protocol for individual templates (MAPit[101]), followed by nucleosome occupancy and methylome sequencing (NOMe-seq)[41,102], which both use an m^{5}C methyltransferase that modifies cytosines in a GpC context.

As genomes of many eukaryotes contain abundant endogenous CpG methylation, and bisulfite sequencing measures methylation on cytosines, exogenous enzymes are required that methylate other dinucleotide contexts. The approach has limited spatial resolution, as it relies on GpC nucleotides that are rare in mammalian genomes, only found once every 20–30 bp, and it is common to find much larger stretches of sequence having no informative positions at all[103]. 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. This method is known as dual-enzyme single-molecule footprinting[104]. This approach has proven to be very powerful in enumerating the distinct functional states of individual promoters, down to the ability to footprint the occupancy of individual components of the basal transcriptional machinery. The approach has also been recently extended to mammalian genomes with sufficient resolution to quantify the single-molecule occupancy patterns of individual TFs at regulatory regions[105]. This requires knock out of endogenous methyltransferases and is limited to the fraction of regulatory regions (typically 30–50%) that contain enough informative GpC dinucleotides. Moreover, bisulfite sequencing-based methods only provide information about the state of individual molecules within, at most, 600-bp stretches of DNA, which is the current limit of combined fully sequenced paired-end read length for Illumina sequencing.

The limited length of the single-molecule read-out obtained via Illumina sequencing reads has been addressed by the advent of long-read sequencing platforms such as PacBio and Oxford Nanopore[106]. In addition to generating multikilobase reads (current record of 2.3 Mbp)[107], these technologies are capable of reading modified bases directly within individual molecules, although with significantly decreased accuracy[108–110]. Base modification detection by long-read sequencing remains challenging, as it may require high coverage as well as training and control data sets to reduce erroneous calls[111]. The accuracy can be increased by using

PacBio circular consensus sequencing, although this reduces the effective read length as there is trade-off between the number of sequencing passes and insert sizes[112]. nanoNOMe-seq and methyltransferase treatment followed by single-molecule long-read sequencing (MeSMLR–seq) assays use GpC methylation and nanopore sequencing to map accessibility on a multikilobase scale, although they are still limited in resolution by available informative positions[113,114].

The limit in the number of informative positions can be overcome by taking advantage of the ability of long-read platforms to read any modification, not just methylated cytosines. For instance, non-specific methylation, such as m^{6}A deposited via EcoGII, or other modifications (Tet-assisted pyridine borane sequencing (lrTAPS)[115]) can be combined with nanopore or PacBio sequencing to obtain a fine-scale read-out of chromatin accessibility at the single-molecule level.

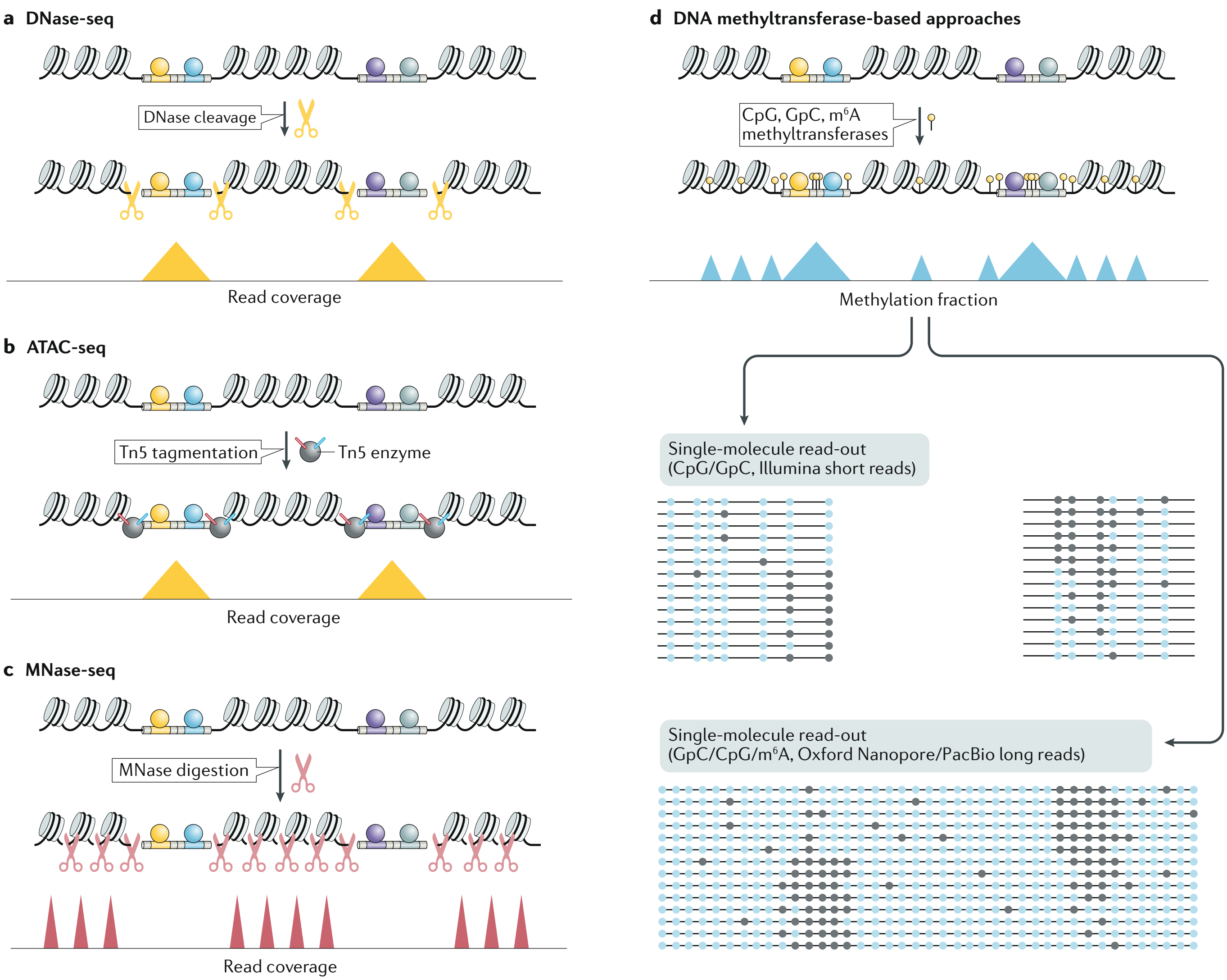

Fig. 2 | **Experimental approaches for measuring chromatin accessibility and nucleosome positioning. a** | In deoxyribonuclease I (DNase I) hypersensitive site sequencing (DNase-seq), the DNase I enzyme (represented as yellow scissors) is used to preferentially cleave accessible chromatin, generating fragments that can then be amplified into sequencing libraries. **b** | In Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), a hyperactive version of the Tn5 transposase (represented by the dark grey circle) is used to preferentially insert into accessible chromatin while simultaneously attaching adapters (represented by the red and blue lines on the Tn5 transposase) to the resulting fragments that can be used to directly amplify sequencing libraries. Both DNase-seq and ATAC-seq generate peaks in read coverage over accessible regions in the genome. **c** | In micrococcal nuclease sequencing (MNase-seq), the MNase enzyme (represented as red scissors) is used to digest DNA that is not protected by bound proteins, leaving intact fragments protected by protein occupancy (primarily nucleosomes). These fragments are then amplified, resulting in increased read coverage over positioned nucleosomes. **d** | DNA methyltransferase-based approaches rely on the labelling of accessible DNA with DNA methylation modifications (represented by drawing pins), which can either be sequenced using Illumina platforms following bisulfite conversion or via long-read sequencing platforms that directly read the modified bases (unmodified and modified bases are represented as light and dark blue circles, respectively). These single-molecule chromatin accessibility profiling approaches tend to provide a simultaneous read-out of both nucleosome positioning and accessible chromatin regions. Accessible chromatin regions represent themselves as higher peaks due the fact that they have more nucleotides that are accessible to the methyltransferases and are therefore more frequently methylated, compared with the internucleosomal sequences that are thus not methylated in every single-molecule read. In all four panels, bound transcription factors (TFs) are visualized via coloured circles on the accessible chromatin.

This can be done either on total genomic DNA — with the single-molecule long-read accessible chromatin mapping sequencing (SMAC-seq) assay[103] or mapping chromatin fibres onto a DNA template using methyltransferases (Fiber-seq)[116] — or in combination with a phasing MNase digestion step (single-molecule adenine methylated oligonucleosome sequencing assay (SAMOSA)[117]). The large number of informative positions allows for fine-scale footprinting almost everywhere in the genome. Although the higher error rate in base calling for long-read sequencing technologies does not yet allow nucleotide resolution for these assays, in practice, the biologically relevant scale of chromatin accessibility typically is larger than that of an individual base. Due to high error rates in the calling of the nucleotides in a read, obtaining a fully correct single-nucleotide read-out is still a challenge. For instance, if 1 in every 20 bp is wrongly identified, then multiple modified nucleotides are taken together to obtain an estimate of the accessibility of that part of the DNA, thereby compromising on the resolution. Nevertheless, the resolution is much higher compared with ATAC-seq, for instance.

ChIP–seq. ChIP–seq is used to detect the occupancy of chromatin-binding factors (such as TFs) or histone modifications at a genome-wide level[34,118–120]. The amino-terminal tails of core histones are enriched with various covalent modifications — including methylation, phosphorylation, acetylation, ubiquitylation and sumoylation — that serve as the docking sites for many chromatin-binding proteins[121,122]. Typical histone marks used to define regulatory elements include histone H3 acetylated at lysine 27 (H3K27ac), which correlates with DNase-seq and ATAC-seq data at transcription start sites, active promoters and distal active enhancers[123,124]; H3 dimethylated at lysine 4 (H3K4me2), which has a similar genomic distribution to H3K27ac; and H3 monomethylated at lysine 4 (H3K4me1), which correlates with poised or active enhancer regions in animals[125,126] when it co-occurs with H3K27me3 or H2K27ac, respectively[126–128].

For ChIP–seq analysis of chromatin modifications, chromatin can be isolated from either formaldehyde-fixed cells or non-fixed cells (native chromatin), and fragmented to 100–500 bp by sonication or through MNase digestion to profile histone modifications[129–131]. For profiling of protein-bound chromatin, for example to determine TF occupancy, the chromatin is cross-linked to stabilize protein–chromatin interactions. Through the use of 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. Chromatin is then reverse cross-linked and the DNA fragments are eluted, end-repaired, ligated to adaptors, PCR-amplified and sequenced by NGS.

Traditionally, ChIP–seq requires at least hundreds of thousands of cells for profiling histone modifications and millions of cells for profiling TFs. ChIP–seq data quality critically depends on antibody specificity, efficiency of chromatin fixation and residence time of the TF on DNA. Each antibody should therefore be screened for ChIP efficiency, and the fixation and sonication conditions need to be optimized for different cell types. The entire procedure for ChIP–seq is time-consuming (spanning multiple days) and laborious.

In the past decade, several ChIP–seq derivatives have been developed involving a lower cell input, detection of TF binding at higher resolution and/or providing a streamlined workflow. These derivatives include ULI-NChIP–seq[132], μChIP–seq[133], small-scale ChIP–seq[134], STAR ChIP–seq[135], ChIPmentation[136], ACT-seq[137], ChIL–seq[138] and CUT&RUN[139]. ChIPmentation combines aspects of ChIP–seq and ATAC-seq by performing tagmentation on immunoprecipitated chromatin fragments, which reduces the input requirement and leads to a simpler, faster assay[136]. CUT&RUN combines antibody-tagging with MNase cleavage in a simple, robust and less expensive protocol for high-resolution profiling of chromatin binding[139]. Recently, some of the above-mentioned methods and other ChIP-derived techniques have even been applied at single-cell resolution, including iACT-seq[137], ChIL–seq[138], scCUT&Tag[140], scChIC-seq[141], CoBATCH[142], uliCUT&RUN[143], Drop-ChIP[144] and scChIP–seq[145].

Single-cell chromatin accessibility

Innovations in barcoding and microfluidics have recently enabled high-throughput biochemical profiling of chromatin accessibility at single-cell resolution, including scDNase-seq[58], single-cell MNase-seq (scMNase-seq[96]) and scATAC-seq[146–150]. 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 focus on the multiple experimental implementations 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 types of scATAC-seq, as well as consideration for experimental design (BOX 2), are described below.

Microfluidics-based scATAC-seq. Droplet-based single-cell partitioning via microfluidic devices has emerged as a powerful approach for single-cell data generation owing to its reproducibility 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)[150] and BioRad (SureCell ATAC-seq Library Preparation Kit)[149], provide all of the reagents needed to produce scATAC-seq libraries. However, these commercial applications require the acquisition of proprietary robotic sample processing devices (Chromium Controller, 10x Genomics; ddSEQ single-cell isolator, BioRad) that are non-standard in most laboratories.

Droplet microfluidic-based scATAC-seq methods generally start by performing Tn5 adapter integration on a bulk nuclei suspension, while keeping the nuclei intact, similar to traditional ATAC-seq. Transposed nuclei are then loaded onto an aqueous channel with PCR reagents

Box 2 | **Experimental design for scATAC-seq**

Similar to other chromatin profiling methods, single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq) is susceptible to batch effects that can obscure biological variation. Careful attention to experimental design is central to mitigating batch effects and other sources of technical variation, but this depends on the goals of the experiment[248,387]. For example, in atlas mapping and in case–control studies, a common objective is to contrast regulatory patterns within and between cell types found in different tissues and organs, or between treatments and control samples. To allow for robust statistical tests on such contrasts, the inclusion of at least two biological replicates is highly recommended. Replicates are useful to identify failed runs and batch effects, and to increase cell counts. Indeed, independent scATAC-seq experiments are often done under the same condition, and subsequently computationally combined to increase the number of cells in the final data set, for instance to yield more power to distinguish small cell populations. 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. By 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[248]. 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.

and suitable buffers and mixed with gel beads containing distinct barcodes. To encapsulate individual nuclei in picolitre reaction compartments with a single gel bead, the aqueous flow is restricted to channels measuring ~55 μm in width[150]. Droplets are produced by exposing the aqueous flow to a continuous stream of oil. Nuclei droplet loading follows a Poisson distribution, and nuclei are loaded at low concentrations. Barcoded sequences with P5 adapters and tail sequences complementary to Tn5-inserted adapters are released from gel beads following droplet generation, enabling PCR amplification and barcoding of accessible chromatin fragments within each droplet in isolation. Finally, the droplet emulsion is broken, and the fragments are purified with magnetic beads and subjected to bulk PCR to attach sequencing indices and P7 sequences[149,150].

Plate-based scATAC-seq. An alternative to the microfluidics approach is to physically separate cells into the wells of plates. Straightforward 96-well and 384-well scATAC-seq protocols have been published[147], but 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 nanolitre wells, known as μATAC-seq, increased the throughput of the assay to a few thousand cells[151].

Combinatorial indexing (sciATAC-seq). Higher throughput can be achieved using a combinatorial indexing strategy, as implemented in single-cell combinatorial indexing ATAC-seq (sciATAC-seq)[55,56,148]. In contrast to microfluidic approaches, sciATAC-seq can be performed with access to standard instruments and reagents (for example, 96-well or 384-well plates, flow cytometry, Nextera TF buffer and so on). Whereas earlier versions of sciATAC-seq require custom-made Tn5, the latest version has been adapted to work with commercially available Tn5 (REF.[152]). The core idea behind combinatorial indexing is the repeated pooling and splitting of cells or nuclei coupled with labelling 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 are also possible, using the 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 those from 10x Genomics or BioRad, in the form of droplet combinatorial indexing or droplet-based single-cell combinatorial indexing for ATAC-seq (dsciATAC-seq)[149].

Combinatorial indexing
A technique that uniquely labels a large number of single molecules or single cells by split-pool barcoding of nucleic acids.

Doublets
Artefactual libraries generated from two cells in single-cell omics experiments. For instance, in droplet-based methods, doublets arise if two cells are captured in a single droplet.

Results

In general, a chromatin accessibility analysis workflow consists of three main steps: preprocessing, peak calling and 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 scATAC-seq data. We will discuss each of the steps in more detail and mention commonly used bioinformatics tools (Supplementary Table 1). Although there is not yet a gold standard in the field, we will mention parts of some general pipelines, such as the ENCODE ATAC-seq Data Standards and Processing Pipeline[156], and we propose specific tools and a guided workflow for analysis of chromatin accessibility data. In this section, we focus on bioinformatic tools that are used for the analysis of the three most commonly used chromatin profiling methods: bulk ATAC-seq, DNase-seq and MNase-seq — we will not discuss the analysis pipeline for the methods based on single-molecule chromatin accessibility profiling as these are still in their infancy.

Preprocessing

As with most high-throughput sequencing data (FIG. 3a), pre-alignment quality control is recommended for chromatin accessibility data and can be performed using FastQC, which produces an HTML to examine sequencing quality, GC bias and over-represented sequences (FIG. 3b). The FastQC report is also produced by MultiQC[157], which includes visualization of further processing steps such as the mapping percentage, alignment scores and number of reads passing filtering. Next, sequencing adaptors should be removed using tools such as *cutadapt*[158], *trimmomatic*[159] and *fastq-mcf*[160], which require the input of known Illumina adaptor sequences. For certain experimental techniques or computational goals (for instance, for MNase-seq data and footprinting analysis in DNase-seq data), and given that sequencing

BAM file
An alignment file format that is the compressed binary version of a SAM file, used to represent aligned sequences.

Irreproducible discovery rate
(IDR). A measure of consistency between biological replicates of high-throughput sequencing experiments. Also used to determine highly stable peak calling thresholds based on reproducibility.

Genomic intervals
Consecutive stretches on a genomic sequence, specified as a chromosomal location range or as a cytoband designation.

Fraction of reads in called peaks
(FRiP score). The fraction of all mapped reads that fall into the called peak regions.

Signal proportion of tags
(SPOT score). The fraction of reads that fall in tag-enriched regions identified using the Hotspot algorithm.

Signature
A set of peaks that is differentially accessible between studied samples and can be used to define a studied cell type or state.

Differential peak calling
A process in which peaks with significantly differentially accessibility between samples are identified.

was done in a paired-end fashion, selecting reads with a desired read length — also referred to as computational size selection — is recommended at this point. For instance, removal of multi-nucleosomal reads is advised for MNase-seq data. This is based on the size of the paired-end reads, as mononucleosomal reads should be ~147 ± 30 bp in length[161]. For DNase-seq, as well as removing multi-nucleosomal reads, an additional in silico filtering step for fragment inserts between 50 and 100 bp for TF binding site detection can be performed, along with the gel-based or solid-phase reversible immobilization-based experimental size selection[74,77]. Trimmed and filtered reads are mapped, or aligned, to an organism-specific reference genome, generating an alignment file (represented in a BAM file format). The most widely used aligners for chromatin accessibility data are *Bowtie2* (REF.[162]) (used in the ENCODE ATAC-seq pipeline[156]) and *bwa-mem*[163] (used in the Cell Ranger ATAC Algorithm) (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[164] and mitochondrial reads. This is particularly important for ATAC-seq data in which mitochondrial reads can make up as high as 75% of the total amount of mapped reads when using the original protocol[36] (FIG. 3d).

A post-alignment quality control step is recommended at this point, involving visualizing accumulated read abundance around transcription start sites, which are generally highly accessible[165] (FIG. 3e). Other quality control metrics include the number of reads, mapping percentages, duplication percentages and visualization of nucleosome patterning via a fragment size distribution plot[50]. Such diagnostic plots can, for instance, be generated using the package *ATACseqQC*[165]. In addition, visually inspecting the distribution of reads across the genome using genome browsers such as IGV[166], UCSC[167], Ensembl[168] or JBrowse[169,170] can further increase insight into the quality of the samples (FIG. 3e).

Peak calling

Following initial read processing and quality control comes one of the crucial steps in chromatin accessibility data analysis, namely defining 'peaks', which are genomic regions with a high accumulation of reads compared with the background (FIG. 3f). These peaks form the basis for most of the downstream analyses. The most widely used tool for peak calling is MACS2 (REF.[171]), 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. It implements a dynamic Poisson distribution to capture local background biases in the genome and to effectively detect peaks[171]. As MACS2 was originally designed for ChIP–seq data, specific parameters (for example, --nomodel) need to be used for peak calling in ATAC-seq or DNase-seq data. The ENCODE ATAC-seq pipeline contains more detailed information on the parameters. Other general and method-specific peak callers exist, for example, ZINBA[172] (general), HMMRATAC[173] and Genrich[174] (ATAC-seq), and F-seq[175] and Hotspot[176] (DNase-seq and ATAC-seq). The signal threshold, which influences the sensitivity and specificity of peak retrieval, is an important parameter to consider during the peak calling step. The default minimum false discovery rate cut-off of 0.05 for MACS2 has been shown to be optimal for a range of DNase-seq data sets[177].

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. The IDR values can also help to define an independent threshold for peak calling[178]. Specifically, following a lenient peak calling with, for instance, MACS2, a core set of IDR peaks can be defined by only retaining peaks that pass a set IDR threshold, such as, for example, 5% (REF.[179]).

As data sets often comprise different samples, the construction of a common set of features, or genomic intervals, is crucial in order to be able to compare samples with each other in downstream steps. Usually, a consensus peak file is used for this purpose, which comprises the set of peaks that are shared between samples, and in which the start and end location of overlapping peaks are adjusted (through the so-called merging of peaks) to thus yield one consensus peak. The ENCODE pipeline provides a workflow with merge and filter steps for constructing a consensus peak file[156], although other tools can serve the same purpose (for example, *consensusSeekeR*[180]). Alternatively, a predefined set of regions or a binned genome can be used as features in downstream analyses[55,64]. For human and mouse studies, the ENCODE SCREEN regions[181] provide comprehensive sets of intervals, as well as two recently published catalogues 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[64]. Although these compendia of accessible regions cover a large fraction of the regulatory genome, some condition-specific accessible regions could potentially be missing.

Finally, an important quality control step is the calculation of the signal to noise ratio, which can done by calculating the fraction of reads in called peaks (FRiP score). For ATAC-seq, the FRiP score 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, which signifies that 40% of mapped reads are located within DHSs[128,156]. These metrics vary depending on the organism, and they can be dependent on the size and complexity of the genome.

Downstream analysis

Usually, chromatin accessibility profiling is performed on multiple samples, comparing treatment versus control, multiple tissues or 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 (REF.[171]), in which mapped BAM files representing treatment and control samples are provided,

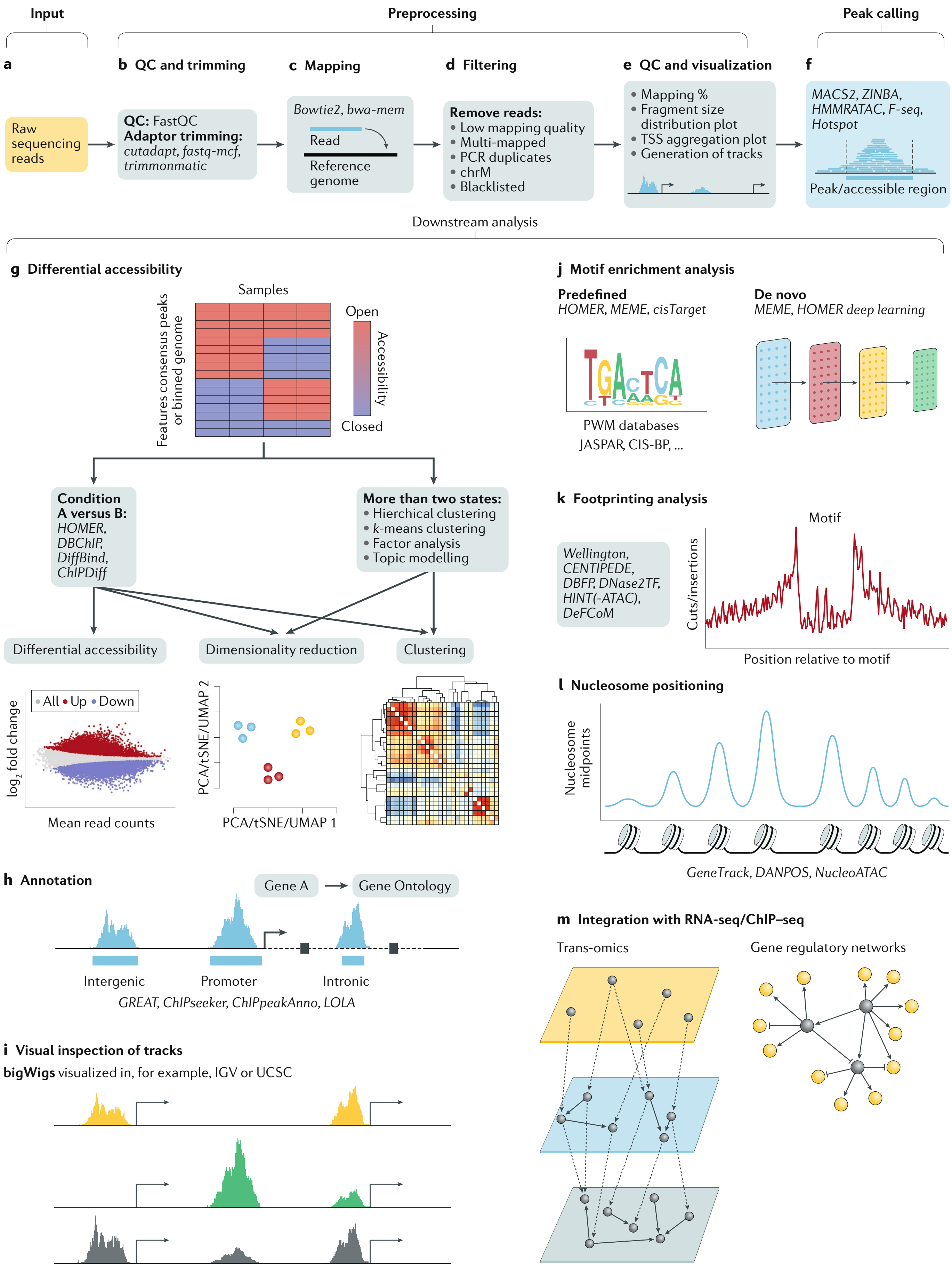
Input
Preprocessing
Peak calling
a
b QC and trimming
c Mapping
d Filtering
e QC and visualization
f
Raw sequencing reads
QC: FastQC
Adaptor trimming: cutadapt, fastq-mcf, trimmonmatic
Bowtie2, bwa-mem
Read
Reference genome
Remove reads:
• Low mapping quality
• Multi-mapped
• PCR duplicates
• chrM
• Blacklisted
• Mapping %
• Fragment size distribution plot
• TSS aggregation plot
• Generation of tracks
MACS2, ZINBA, HMMRATAC, F-seq, Hotspot
Peak/accessible region
Downstream analysis
g Differential accessibility
Samples
Features consensus peaks or binned genome
Open
Accessibility
Closed
Condition A versus B: HOMER, DBChIP, DiffBind, ChIPDiff
More than two states:
• Hierchical clustering
• k-means clustering
• Factor analysis
• Topic modelling
Differential accessibility
Dimensionality reduction
Clustering
All Up Down
log2 fold change
Mean read counts
PCA/tSNE/UMAP 2
PCA/tSNE/UMAP 1
h Annotation
Gene A
Gene Ontology
Intergenic
Promoter
Intronic
GREAT, ChIPseeker, ChIPpeakAnno, LOLA
i Visual inspection of tracks
bigWigs visualized in, for example, IGV or UCSC
j Motif enrichment analysis
Predefined
HOMER, MEME, cisTarget
De novo
MEME, HOMER deep learning
PWM databases
JASPAR, CIS-BP, ...
k Footprinting analysis
Wellington, CENTIPEDE, DBFP, DNase2TF, HINT(-ATAC), DeFCoM
Motif
Cuts/insertions
Position relative to motif
l Nucleosome positioning
Nucleosome midpoints
GeneTrack, DANPOS, NucleoATAC
m Integration with RNA-seq/ChIP–seq
Trans-omics
Gene regulatory networks

Fig. 3 | **Overview of common tasks in the analysis of bulk chromatin accessibility data. a** | Starting from raw sequencing reads. **b** | Bulk chromatin accessibility data generally undergo several preprocessing steps, including a pre-mapping quality control (QC) and adaptor trimming step. **c** | Mapping of the trimmed reads to a reference genome for the studied species. **d** | Mapped reads are filtered. **e** | An additional post-alignment QC step is recommended through several metrics and data visualizations. **f** | An important step in chromatin accessibility data analysis is calling peaks, as these usually form the basis of several downstream analyses. **g** | Differentially accessibility analysis can be performed pairwise (condition A versus B) or across multiple conditions. **h–m** | Additional downstream analyses include annotation and enrichment analysis for the identified peaks (part **h**), visual inspection of chromatin accessibility data tracks (part **i**), motif enrichment within peaks (sets) using predefined databases or de novo (part **j**), transcription factor (TF) footprinting analysis (part **k**), mapping of nucleosome positions (part **l**) and integration with RNA sequencing (RNA-seq) or chromatin immunoprecipitation followed by sequencing (ChIP–seq) data to link different omics layers or to generate gene regulatory networks (part **m**). TSS, transcription start site.

and for which biological replicates are combined prior to differential accessibility analysis. Alternatively, statistical analyses can be performed on the count matrix, a data table containing the number of reads per feature across the samples, yielding a matrix with features as rows and the different samples as columns. As explained above, these features can be the set of consensus peaks, a predefined set of regions or a binned genome. 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, which are implemented in the *DESeq2* (REF.[182]) and *edgeR*[183] packages or in more chromatin accessibility-specific tools such as *DiffBind*[184], *HOMER*[185] or *DBChIP*[186] (FIG. 3g). Data normalization is recommended when comparing conditions or tissues, as library-specific biases or global chromatin accessibility differences can affect differential accessibility results. Multiple data normalization methods exist, for instance normalization for library size or for FRiP scores, quantile normalization and trimmed means of *M* value normalization[187]. Often, the differential accessibility tools mentioned above already include one of these normalization methods[188]. The choice of data normalization method can alter differential accessibility results. A typical way to visually assess the effectiveness of normalization is through an MA plot: a single cloud should be present and the MA distribution should not display an upward or downward shift[188,189].

MA plots
Visual representations of genomic data used to compare two samples or two groups of samples. The x axis represents the base mean value of the samples and the y axis the difference between them.

Hierarchical clustering and k-means clustering
Clustering algorithms that group similar objects in a data set into groups called clusters. In k-means clustering, the data are divided into a predefined number ('k') of clusters, whereas in hierarchical clustering, a hierarchy of clusters is built without requiring a predefined number of clusters.

For differential accessibility analysis in multi-sample studies, several options are possible. One way is to use the normalized count matrix for dimensionality reduction and clustering, for example by hierarchical clustering and k-means clustering (FIG. 3g). Clustering allows one to group together samples with similar chromatin accessibility profiles, as well as to distinguish sets of regions that are differentially accessible across the samples and to generate groups of co-accessible regions (meaning regions that show concordant accessibility patterns across the different samples). Such clustering algorithms are, for instance, implemented in the *deepTools* package[190] and can be visualized with a heat map (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 in order to identify differentially accessible regions. For instance, topic modelling or non-negative matrix factorization, in which a high-dimensional data set is approximated by a reduced number of representative components, can be applied directly to bulk data sets, or to a matrix with simulated single cells, created from bulk samples using a bootstrapping procedure[191,192].

To gain biological insight into the sets of cell type-specific regions identified via differential accessibility analysis, region set enrichment analysis using *GREAT*[193], *ChIPseeker*[194], *ChIPpeakAnno*[195], *Enrichr*[196], *cisTarget*[197,198], *GIGGLE*[199] and *LOLA*[200] is used to identify correlations of peak sets with genome annotation (for example, promoter, intronic, intergenic) or with existing ChIP–seq data sets and to couple peaks to the nearest gene, followed by Gene Ontology or pathway enrichment (FIG. 3h). In addition, chromatin segmentation approaches such as *ChromHMM*[201], *EpicSeg*[202] and *Segway*[203] are used for genome-wide classification of genomic regions into chromatin states (such as 'active promoter' or 'weak/poised enhancer') based on epigenomic marks. These enrichment analyses and genome or chromatin state annotations can be useful in interpreting gained or lost accessible regions in a study. Finally, the generation of tracks — which are a way to display data per sample across the genome, specifically, here, chromatin accessibility data — and visually inspecting them together with such annotations or other public ChIP–seq or RNA-seq data can help to gain further biological insights, such as providing indications on the putative functionality of accessible regions or of the TFs bound to them (FIG. 3i).

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. As many TFs recognize and bind to TF-specific DNA sequences, we can leverage the enrichment of TF motifs in a set of sequences (FIG. 3j). Two major classes of motif analysis tools exist. The first class includes *HOMER*[185], *MEME*[204] and *cisTarget*[197,198], and relies on databases of predefined TF motifs (position weight matrices[205]) such as *JASPAR*[206], *CIS-BP*[207], *TRANSFAC*[208] and *HOCOMOCO*[209]. These approaches scan the DNA sequences of accessible regions with position weight matrices and perform an enrichment analysis compared either with a background set or with the entire genome as background. The second class of tools includes *RSAT*[210], *MEME*[204], *Weeder*[211] and *HOMER*[185], and performs de novo motif discovery, allowing unsupervised identification of enriched TF motifs. Although the identification of de novo motifs does not require a motif database, motif databases are still needed to link de novo motifs to known TFs.

Recently, machine learning methods such as the convolutional neural network models *Basset*[212], *DeepSea*[213], *DeepLIFT*[214] and *DeepMEL*[215] have shown promising results to predict TF motifs in accessible regions in a more precise and unbiased manner. Such models are trained on large sets of co-accessible peaks per cell type, can capture important TF motifs across the training regions and are able to predict their importance at single-nucleotide resolution within the regulatory sequences. Altogether, motif detection on a set of specifically accessible regulatory regions allows the decoding of 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. 3k). TF footprints are small regions (8–30 bp) that display relative protection from cleavage due to binding of a TF and thus correspond to dips in the accessibility peak[47,216,217]. DNase I has been, and is still, the preferred footprinting reagent. Analytic genomic footprinting approaches such as the Wellington algorithm[218], *HINT*[219], *DBFP*[220] and *DNase2TF* (REF.[221]) de novo annotate DNase I footprints, or they determine TF occupancy at a specific genomic location, such as *CENTIPEDE*[222] and the footprint likelihood ratio[223,224]. 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 for human samples[224]. In addition, TF footprinting is biased by the short residence times of some TFs on DNA and by intrinsic sequence preferences of the cleavage enzymes, which should be corrected for[78]. In general, ATAC-seq footprinting has been shown to be less accurate than DNase-seq footprinting[225], 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[36,226]. Nevertheless, footprinting analysis on ATAC-seq data has been performed with success by several groups, for instance in the initial ATAC-seq publication[36], using *DeFCoM*[227] or by using ATAC-seq-specific footprinting algorithms (such as *HINT-ATAC*[226] and *TOBIAS*[228]) that consider ATAC-seq artefacts and, for instance, correct for Tn5 transposase cleavage biases.

MNase-seq is orthogonal compared with 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 (for example, *DANPOS*[229]) have been developed[229,230] (FIG. 3l). Note that similar to TF footprinting analysis, correction for enzymatic cleavage bias should be performed in nucleosome footprinting analysis. ATAC-seq also lends itself to nucleosome positioning by partitioning paired-end reads based on their size to separate putative nucleosome free reads and mononucleosomal, dinucleosomal and trinucleosomal reads[36] or by using specific tools such as *NucleoATAC*[231]. In addition, Zhong et al. have shown that DNase-seq data can also be used to infer nucleosome positions with high accuracy by using a Bayes factor-based nucleosome scoring method[232]. Therefore, all three commonly used chromatin accessibility profiling methods lend themselves to detect nucleosome positioning genome-wide (provided that the data are sequenced paired-end), although MNase-seq is still the most frequently method for this purpose.

Pseudo-time trajectory
A computational reconstructed path of a dynamic biological process, such as differentiation, undergone by the cells in a single-cell omics experiment. Single cells are ordered along the trajectory based on their 'pseudo-time', or their inferred progression through the biological process.

Single-cell data analysis

Single-cell chromatin accessibility data require similar upstream processing steps to 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[191,233–242]. One major point in which these tools differ is the way they define genomic regions to be used as features, either as peaks from bulk or aggregated single-cell data (*chromVar*[239], *Cicero*[238], *cisTopic*[191], *scABC*[241], *Scasat*[233], *MAESTRO*[242]), peaks from pseudo-bulk samples[56] or fixed-size bins[56] (*SnapATAC*[243]). Another difference between the bulk and single cell-based algorithms is what the count features represent, for example, counting reads in peaks (*cisTopic*[56,191], *scABC*[241], *Scasat*[233], *MAESTRO*[242]), counting gapped *k*-mers under peaks or around transposase cut sites (*BROCKMAN*[234], *chromVAR*[239]), or counting reads overlapping TF motifs in peaks or genome-wide (*chromVar*[239], *SCRAT*[237])[244]. *ArchR*[236] uses an iterative feature definition method; it first defines a feature-by-cell count matrix of the number of reads per feature (in this case, 500-bp genomic bins) across all single cells, which then undergoes an iterative latent semantic indexing reduction to generate the cell clusters and pseudo-bulk samples on which peaks are called.

Important follow-up steps are transformation and dimensionality reduction of the feature by cell count matrix to visualize the cells in a 2D or 3D space and performing further downstream analyses, for example 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 pseudo-bulk BAM files and tracks to visualize the data (FIG. 4d). Recently, ten computational methods for the analysis of scATAC-seq data have been benchmarked[244], demonstrating that *SnapATAC*[243], the method used in REF.[56] and *cisTopic*[191] performed best in distinguishing cell populations. Next to these, extensions of popular scRNA-seq analysis toolkits specifically designed for single-cell chromatin accessibility data, including *Signac* (an extension to *Seurat*[245]) and *EpiScanpy*[246] (an extension to *Scanpy*[247]), are used in the field.

There are currently no designated tools that correct for batch effects in scATAC-seq data. Inexplicit batch correction is performed during the processing steps such as during feature selection or dimensionality reduction[248]. Batch correction tools designed for scRNA-seq data[249–252] may be used with precautions not to remove biological variance, for instance by assessing the retainment of biological variation between easily defined cell labels and cell trajectories[253]. To avoid overcorrection, it is recommended to compare multiple batch correction methods to obtain the best result for a given data set. Batch correction becomes especially important when combining multiple runs into atlases or when integrating scRNA-seq data, for which *BBKNN*[249], *Scanorama*[250] and *scVI*[251] performed best in a recent benchmark[253]. As in scRNA data, reconstruction of a pseudo-time trajectory based on scATAC-seq data can be helpful when studying a system following a cellular differentiation, for instance during embryonic development[254] or haematopoiesis[255] (FIGS 4f,g). Tools such as *Cicero*[238] (which implements modified aspects of the scRNA-seq trajectory inference tool *Monocle*[256], and *STREAM*[255]) have been used to infer such trajectories from scATAC-seq data.

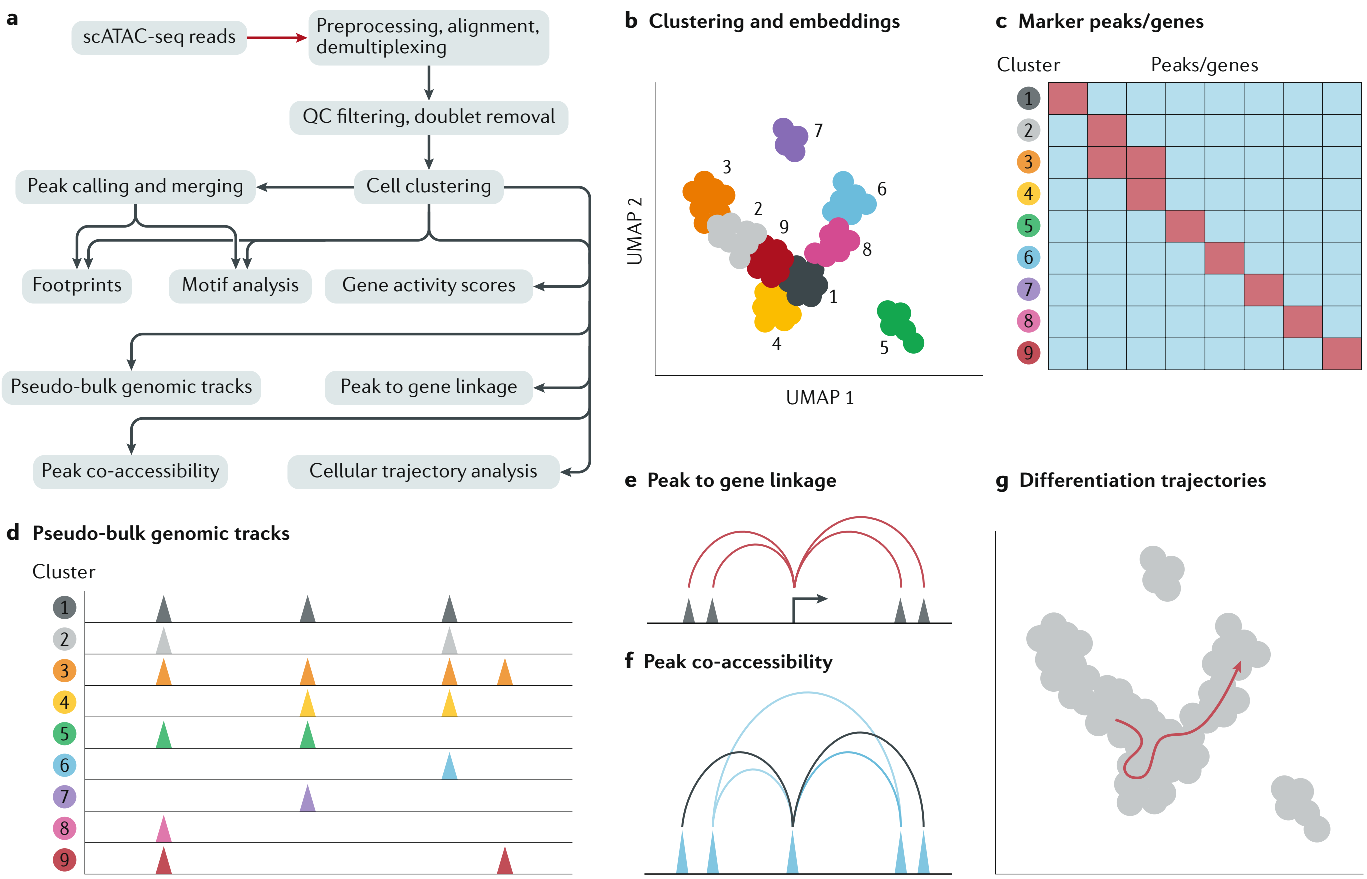

Fig. 4 | **Overview of common tasks in the analysis of scATAC-seq data. a** | Outline of key steps in processing single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq) data sets, six of which are illustrated in panels **b**–**g**. **b** | An important step in the analysis of scATAC-seq data is clustering the cells via dimensionality reduction of the feature by cell matrix (via UMAP, for example) to discover the different cell populations. In the given example, dots represent the single cells, and their colours and numbers represent the nine identified cell clusters or cell populations. **c** | Identification of marker genes and/or peaks for each of the cell clusters allows further study of the putative cell populations. **d** | By aggregating the accessibility profiles of all cells within a cluster, pseudo-bulk genome browser tracks can be generated for each cell population. **e** | Specific tools allow the identification of peak to gene links, which can reveal putative target genes of identified peaks. **f** | Assessing peak co-accessibility allows grouping peaks into sets of co-regulated regions. **g** | When analysing scATAC-seq from a time-series or differentiation experiment, trajectory analysis allows study of the dynamic changes in chromatin accessibility along a pseudo-time axis. QC, quality control.

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. The integration of chromatin accessibility profiles with ChIP–seq and RNA-seq data is of particular interest for inferring the binding of specific TFs and for reconstruction of regulatory networks (FIG. 3m). The integration of epigenomics and transcriptomics may predict links between accessible regulatory regions and target genes (FIG. 4e). An example from the single-cell field involves the use of a least absolute shrinkage and selection operator (LASSO) model to correlate a gene's expression level with the accessibility of all peaks within 100 kbp around its transcription start site, linking 1,260 distal regions to 321 potential target genes[257]. This improved the prediction of gene expression based on accessibility profiles fourfold compared with only using chromatin accessibility at promoters.

Applications

Chromatin accessibility profiling is widely useful for applications in biology and biomedicine, ranging from the analysis of gene regulation and cellular states to the dissection of healthy and diseased tissues and organs, and the investigation of populations and species. These applications benefit from the high genomic resolution of chromatin accessibility profiling, from robust and straightforward assays with low input requirements and from the ability to process many samples in a fast and reasonably cost-efficient manner.

Regulation of chromatin accessibility

As nucleosome occupancy of DNA is refractory to TF binding and transcription, regulation of chromatin accessibility is key to gene regulatory mechanisms. Multiple mechanisms for accomplishing chromatin accessibility have been proposed. Nucleosomes appear to have clear preference for certain sequences, and this

Zygotic genome activation
A process by which transcription is turned on after fertilization, making the switch from an unfertilized oocyte with nearly any gene expression to a state where up to thousands of genes are transcribed.

bias seems to play some role in establishing nucleosome positions in yeast[258,259]. However, as this bias is less predictive of nucleosome positioning in metazoan genomes[260,261] and accessible regions are mostly relatively large (that is, hundreds of base pairs) and associated with active *cis*-regulatory elements, sequence preference of nucleosomes is likely not a major contributor to the regulation of chromatin accessibility. Controlling regulatory element accessibility and activity is accomplished through the combined action of TFs, RNA polymerases, chromatin-remodelling complexes, histone chaperones and histone variants[1,262].

Many developmental processes involve chromatin remodelling, especially to make previously inaccessible regions accessible. This process is most noticeable in zygotic genome activation during embryonic development, when transcription of the zygotic genome is turned on. Chromatin remodelling is also important for subsequent lineage-specifying developmental transitions, responses to many external and internal stimuli, and cellular reprogramming. Pioneer factors are particularly important in regulating these processes as they are capable of binding at previously inaccessible chromatin and they subsequently initiate the formation of an accessible state[23]. Well-known examples of pioneer factors include Zelda, which acts upon zygotic genome activation in *Drosophila*[263–265], the Nanog/Oct/Sox pluripotency factors[266–268], FoxA[21] and numerous others[269]. Pioneer factors do not create and maintain an active and accessible state on their own. Rather, they recruit other TFs and chromatin remodellers, and reposition nucleosomes and chromatin modifiers that deposit histone marks characteristic of active regulatory elements[23,269]. Note that general TF binding can form a constraint for the reappearance of nucleosomes in accessible regions.

Cell state transitions also involve the 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, H3K9me3 and, eventually, DNA methylation[270]. This process effectively remodels the chromatin to an inaccessible state.

Cell types and organs

Chromatin accessibility at gene-regulatory regions is highly dynamic during cellular differentiation and organ development[17,271]. Chromatin accessibility profiling has contributed to our understanding of chromatin regulation across a broad range of organs and cell types in human, mouse, *Drosophila* and other model organisms[53,56,128]. The haematopoietic lineage, in particular, has served as a blueprint for deciphering the role of chromatin accessibility and epigenetic changes in cellular differentiation[33,272]. Application of ATAC-seq and/or ChIP–seq to flow cytometry-purified haematopoietic cell populations has established comprehensive maps of regulatory regions and their dynamic changes in the haematopoietic lineage of human and mouse[37,125,273,274]. Detailed investigations of macrophages have connected the regulation of these immune cells to their tissue environment[275,276], whereas analyses of CD4$^+$ T cells[36,277,278] and innate lymphoid cells[279,280] have 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[281] that are poised to actively respond upon re-exposure to pathogens, as well as to the more limited memory of inflammation in regulatory T cells[282]. Importantly, immune cell memory is not restricted to B cells and T cells but also includes monocytes and natural killer cells[283], and the regulation of such trained immunity appears to involve tightly regulated changes in the epigenomes of the affected cells[284,285].

Beyond the haematopoietic lineage, RNA-seq, ATAC-seq and ChIPmentation profiling in epithelial cells, endothelial cells and fibroblasts from many different organs have uncovered widespread immune gene regulation in these structural cells, and an epigenetic potential that appears to preprogramme these cells for contributing to pathogen response[286]. Chromatin accessibility has also been studied in neural development[61,287–289] and in brain samples of humans[57,59,290] and non-human primates[291]. Notable applications of chromatin accessibility profiling to other cell types and organs include the analysis of cardiac development[292,293], epidermal progenitor cells in the skin[294] and mammary gland development[295]. Finally, initial single-cell atlases of chromatin accessibility across tissues and organs are emerging[55,56,60,152], 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 flow cytometry. 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.

Human diseases

Changes in chromatin accessibility are 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[296]. In blood cancers, chromatin accessibility patterns are 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[37,297–300]. Changes in chromatin accessibility have been investigated over the course of targeted therapy in patients with chronic lymphocytic leukaemia[301] and combined with chemosensitivity screening to identify promising drug combination therapies[302]. Chromatin accessibility landscapes have also been mapped in solid tumours, including breast cancer[145], colon cancer[303,304], glioblastoma[305,306], gastric cancer[307] and lung cancer[308,309]. Paediatric 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 tumour cells[310,311]; and epigenome profiling has uncovered subtype-specific regulatory mechanisms in atypical teratoid rhabdoid tumours[312] and in Langerhans cell histiocytosis[313].

Quantitative trait loci
(QTL). Small regions of the genome at which a genetic variant is associated with a quantitative trait of a cell or an organism, based on statistical association between genetic markers and the measurable trait.

Yoruba HapMap
A resource set up by the Yoruba HapMap project that aims to catalogue the common patterns of human genetic variation and associate SNPs with genotypes across human populations.

Hi-C
(High-throughput chromosome conformation capture). A genome-wide sequencing technique used to investigate 3D chromatin conformation.

Chromatin accessibility QTL
Quantitative trait loci (QTL) associated with chromatin accessibility. Specifically, chromatin accessibility QTL represent an SNP that is correlated significantly with accessibility changes in their encompassing region.

An interesting line of research has investigated the role of the tumour-associated immune cells in solid tumours. Regulatory changes are implicated in T cell exhaustion in the context of chronic inflammation and the tumour microenvironment[314,315], which compromises the ability of these T cells to fight the tumour. Immunotherapy, most notably blocking of the PD1/PDL1 pathway, can revert some of the regulatory changes associated with T cell exhaustion[150,316,317] and is widely useful for the treatment of those solid tumours that have a high degree of immunogenicity[318]. However, not all exhausted T cells are rejuvenated by immune checkpoint blockade, as some T cells appear to transition to a fixed regulatory state that renders them resistant to reprogramming[314].

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[319] and rheumatoid arthritis[320]. Changes in epigenome and chromatin accessibility profiles have been observed in post-mortem brain tissue from patients with Alzheimer disease[321], schizophrenia[322] and autism spectrum disorder[323]. 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 as yet received little attention.

Variation within populations

Extension of chromatin accessibility assays to populations of diverse genetic backgrounds is 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 genome-wide association studies, localize to non-coding loci distant from the affected gene, obfuscating functional predictions[29,324,325]. Mounting evidence implicates the alteration of gene regulation as a key driver of phenotypic evolution and disease proliferation. Quantitative trait loci (QTL) mapping of molecular traits, such as gene expression variation (expression QTL), 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[326]. This study also provided evidence that sequence alterations underlying *cis*-regulatory 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 related locus[327]. Thus, the majority of genetic variants located within accessible chromatin appear to lack functional consequences on gene regulation. The same study further demonstrated that pairwise correlations of accessible regions (co-accessible regions) readily recapitulate 3D higher-order chromatin interactions as defined by in situ Hi-C (high-throughput chromosome conformation capture). This similarity suggests that local chromatin accessibility among pairs of regions is coordinated with higher-order genome structure, particularly within the same topologically associated domains. In line with these findings, local chromatin accessibility in a subset of regions has been associated with variants located tens to hundreds of kilobase pairs away, reflecting putative long-distance functional 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 homozygous inbred individuals, making the identification of chromatin accessibility QTL more straightforward. A critical subset of chromatin accessibility QTL could be explained by making or creating binding motifs for pioneer factors[328]. In an alternative approach, chromatin accessibility can also be compared between alleles, within the same individual, to identify allele-specific chromatin accessibility[329].

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. Future 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.

Evolution of chromatin accessibility

Chromatin accessibility profiling facilitates the identification of causal genetic variants underlying disease and trait variation. Similar methods and analyses have also proven 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, are linked to loss of regulatory elements[330]. 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 were used to identify promoters and enhancers in the liver tissue of 20 mammalian species[259]. The rate of sequence variation was much greater for enhancers in comparison with promoters. This was reflected in a lower conservation of enhancers between species, yet newly evolved enhancers are more likely to be under positive selection in a lineage-specific manner.

In plants, incorporating chromatin accessibility data into evolutionary studies helps with the identification of *cis*-regulatory elements, as this identification through sequence-based alignment alone is often hindered by the high proportion of DNA sequence variation in intergenic regions[331,332]. Chromatin accessibility profiling can help reveal important clues about the evolution of gene regulation. For instance, a comparative epigenomics study of numerous flowering plant species, ranging in genome size from ~150 to 5,000 Mbp, revealed rapid evolution of *cis*-regulatory sequences within accessible chromatin regions where DNA sequence conservation was detected[333]. The frequency of distal accessible chromatin regions was correlated with genome size and their distance from genes was mostly due to transposon and repeat expansion in these plants[69,330,334].

The lack of distal regulatory regions in *Capsaspora owczarzaki*, a unicellular eukaryotic organism sister to other animal species, has led to the hypothesis that distal regulation is a feature of animal multicellularity[335]. However, with the increase in profiles of chromatin accessibility across taxa, it seems more likely that distal regulation is a consequence of genome size[333]. 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.

Reproducibility and data deposition

The genomics community has been leading the way in creating standards for data information, data quality and data deposition for decades (TABLE 2). Many genome-wide data sets serve as community resources and, as a result, are repeatedly used and incorporated into future studies by individual laboratories. 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[336] at the National Center for Biotechnology Information (NCBI) or the ArrayExpress database[337] at the European Bioinformatics Institute (EBI). These databases host records of genomics data containing not only count matrices and other useful processed output files (for example, 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 both bulk and single-cell chromatin accessibility data, researchers are expected to submit their FASTQ files to specific databases. In the case of scATAC-seq, specifically for data generated via the 10x Genomics platform, preferably three FASTQ files are submitted, namely the FASTQ file containing the barcode read and the two FASTQ files containing the paired-end feature reads. For non-human species and open-consent human donors, the raw sequencing data should be submitted to, for instance, the

Table 2 | **Commonly used databases for archiving and distributing chromatin accessibility data**

Database type	Database	Description
General epigenomic databases	Gene Expression Omnibus (GEO)	Repository that archives and distributes microarray and high-throughput sequencing data submitted by the research community
	ArrayExpress	Repository that stores and allows sharing of data from high-throughput functional genomics experiments
Databases to deposit raw sequencing data	Sequence Read Archive (SRA)	Largest publicly available repository of high-throughput sequencing data
	European Nucleotide Archive (ENA)	Database for archiving and sharing all types of nucleotide sequencing data
	DNA Data Bank of Japan (DDBJ)	Database for archiving and sharing all types of nucleotide sequencing data
	European Genome–phenome Archive (EGA)	Database for archiving and sharing all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects
	Database of Genotypes and Phenotypes (dbGaP)	Repository for archiving and distributing individual-level human data and results from studies that have investigated the interaction of genotype and phenotype
Databases to deposit code	GitHub	Platform on which researchers can host software development and perform version control using Git
	Zenodo	Repository for the deposition of both code and data
	Kipoi	Repository of ready to use trained machine learning models for genomics
Databases that make processed data easily accessible: portals of large consortia	Encyclopedia of DNA Elements (ENCODE)	Consortium with the goal of building a comprehensive list of functional genomic elements in the human genome using various omics assays
	Roadmap Epigenomics	Consortium aiming to deliver a collection of normal epigenomes (via histone ChIP–seq, DNase-seq, etc.) across a broad range of cell types that can serve as a reference for future studies
	BLUEPRINT	Consortium effort to map epigenomes of the haemopoietic system for healthy and diseased individuals
Databases that make processed data easily accessible: portals based on meta-analyses	ChIP-Atlas	Integrative database for visualizing and making use of public ChIP–seq data
	ReMap	Platform of integrative analysis of *Homo sapiens* and *Arabidopsis thaliana* transcriptional regulators from DNA-binding experiments, including ChIP–seq
Databases that make processed data easily accessible: study-specific portals	Many, e.g. mouse sci-ATAC-seq Atlas	Laboratory-specific, often include several tabs covering, e.g., data visualization and data download

For each of the databases, a short description of the goal or properties is given. sciATAC-seq, single-cell combinatorial indexing for Assay for Transposase-Accessible Chromatin using sequencing; ChiP–seq, chromatin immunoprecipitation followed by sequencing; DNase-seq, deoxyribonuclease I (DNase I) hypersensitive site sequencing.

Sequence Read Archive (SRA)[338], European Nucleotide Archive (ENA)[339] or DNA Data Bank of Japan (DDBJ)[340]. For human donors where controlled access is required for adequate data protection, the raw sequencing data should be submitted, for instance, into the European Genome–phenome Archive (EGA)[341] or the database of Genotypes and Phenotypes (dbGaP)[342] from the NCBI.

To facilitate interpretation and reproducibility, the deposited data should include metadata. For example, data entry requirements that are useful to address issues associated with reproducibility could include sources of possible biological and technical variation. Sources of biological variation include genotype, sex of samples, age and tissue/organ/cell type, whereas sources of technical variation could relate to antibodies (requiring reporting of the lot number) and nucleases/integrases (requiring reporting of the lot number, sequencing library procedure, instrument used for sequencing and type of sequencing run). These possible sources of technical and biological variation are important variables that can be incorporated into data analyses as covariates or to correct for batch effects. The versions of genome assemblies and genome annotations used in data analyses should also be provided.

As well as the mentioned general epigenomic databases (GEO[336] and ArrayExpress[337]), several secondary databases make chromatin accessibility data, or other types of omics data, easily available — often by including an experiment matrix providing a visual representation of the available data, classified based on assay type, tissue type, organism and so on. Such databases include data portals set up by large consortia that produce large numbers of epigenomes (for example, ENCODE[30], Roadmap Epigenomics[32], IHEC[343] and BLUEPRINT[33]), data portals that originate from reanalysis or meta-analysis (ChIP-Atlas[344] and ReMap[345]), UCSC track hubs[346] or interactive web-based tools hosted by individual laboratories. The latter, although useful for initial exploration of the generated data, are often less sustainable and vary in their set-up, and are therefore not the preferred platform to disseminate epigenomics data in a reproducible way. Nevertheless, custom websites often provide integrated downstream analysis results that are context-specific and go beyond standard data formats[56,347]. There exist dedicated tools and websites to host and visualize single-cell chromatin accessibility data, including the use of SCope[348], a Shiny app[237] or ASAP[349]. Specifically for single-cell chromatin accessibility data, we envision that in the near future specific platforms will arise that provide an overview of large amounts of public data and facilitate their easy dissemination, similar to the Single Cell Expression Atlas[350], a data portal hosted by the Human Cell Atlas.

Finally, the distribution of custom code and the documentation of computational methods are also paramount to reproducibility. The ENCODE Consortia has developed extensive open-source software that is accompanied with a document of 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 in more specialized repositories such as Kipoi[351] 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.

Limitations and optimizations

Although chromatin accessibility has proven a powerful and informative window into gene regulation, it is often combined with other measurements and with perturbations to build a causal or mechanistic understanding of genomic function. Whereas accessibility dynamics can be readily profiled, the specific molecular factors that drive accessibility changes may only be inferred by changes in the accessibility or footprints associated with DNA motifs. However, directly inferring the specific TF that is possibly bound based just on the observed DNA motifs is tricky, as a DNA motif may be bound by various related TFs, often within a TF family of structurally similar DNA binding domains. One way of narrowing down the specific TF of a TF family that is bound to DNA motifs enriched in regions undergoing accessibility changes is by determining which of the TFs undergoes concomitant changes in gene expression. Still, to mechanistically link the binding of a specific TF, subsequent experiments are needed, such as TF knockdown or ChIP–seq targeting the specific TF implicated.

Additionally, chromatin accessibility of a putative regulatory locus is usually necessary but not in itself sufficient for bona fide functional regulation. Other marks, 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[352–354]. Therefore, chromatin accessibility data should be combined with other genomic assays to build a stronger set of inferences on the functionality of specific elements.

Finally, chromatin accessibility profiling using DNase-seq and, to a lesser extent, ATAC-seq may require optimization of the reaction time, lysis protocols, cell handling and freezing or thawing, as well as library purification, to produce optimal data. For methods such as ATAC-seq, numerous 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.

Outlook

The current and future challenge in the study of chromatin accessibility is to dissect the function of these regulatory regions in relation to other regulatory layers and gene expression (FIG. 5). Chromatin accessibility alone provides no information on the functional properties of the region — whether it acts as a promoter, enhancer, silencer or replication origin — or its activity state. Information on which TF factors are likely bound to the region must be inferred through sequence analysis.

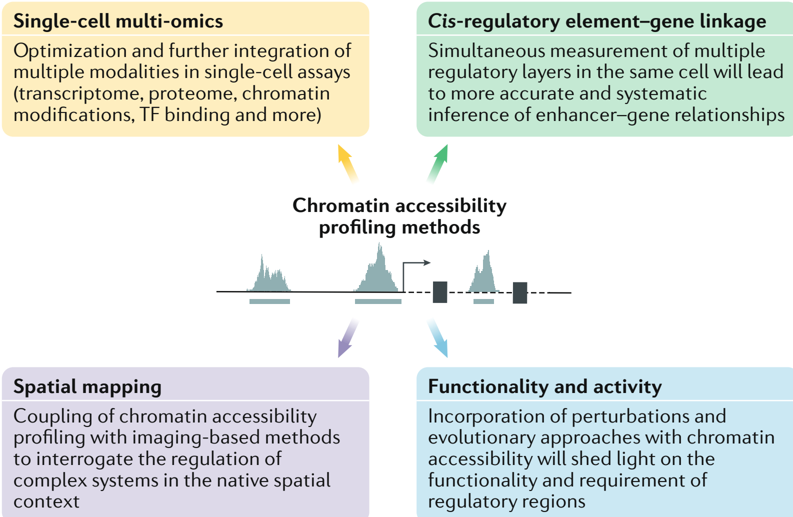

Fig. 5 | **Schematic overview of future roads and opportunities for chromatin accessibility profiling.** In the coming years, our ability to measure chromatin accessibility concurrently with multiple regulatory layers in the same single cell will continue to expand. New insights into regulatory biology 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 data sets will be crucial for the success of these endeavours. Ultimately, these approaches will empower us to functionally dissect the role of regulatory elements and their relationship to gene expression. TF, transcription factor.

Many of these challenges can be overcome by a more holistic multi-omics approach, by profiling the transcriptome, histone modifications and TF occupancy from the same sample, 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 data sets[355,356].

Chromatin accessibility profiling in single cells has surged dramatically in recent years, and we expect further improvements in the coming years as this trend increases. The analysis of accessibility and other regulatory features at the single-cell level is challenging. 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 are mostly binary and very sparse due to the low coverage per cell. A certain degree of data aggregation across cells or features is usually required. 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[191,233–241]. However, tools designed for scATAC-seq for specific analysis tasks, such as pseudo-time and trajectory inference, remain limited. Although comparisons of performance and applicability of scATAC-seq methods have been performed[244], there are no uniform pipelines being widely used by the community, which complicates the systematic comparison and interpretation of results coming from different laboratories. In the coming years, we foresee major efforts in the standardization of comprehensive computational pipelines for the analyses of single-cell epigenomic data. In addition, it is difficult to estimate the sensitivity of scATAC-seq. Roughly, ~10–15% of known peaks are recovered per single cell[146], but it is not known how many regulatory elements are accessible in any given cell at any instance in time. Technical advances have improved cell coverage, which ameliorates both issues and has led to a significant increase in assay sensitivity, allowing a sharper distinction between cell types as well as regulatory changes. Nevertheless, when homogeneous or flow cytometry-purified samples are used as input, then bulk ATAC-seq will likely remain the preferred assay.

Recent advances in single-cell methods are pushing technologies to perform multi-omics measurements simultaneously from the same single cell. Multiple methods have already been published for simultaneous scATAC-seq and transcriptome profiling. These include sci-CAR[257], Paired-seq[155] and SHARE-seq[153], which are all based on combinatorial indexing, as well as SNARE-seq[357] and 10x Genomics Chromium Single Cell Multiome[358], which are droplet-based microfluidics methods. Other achievements include joint profiling of chromatin accessibility with either protein-level quantification (Pi-ATAC[359]) or with DNA methylation (scNOMe-seq[360], COOL-seq[361], EpiMethylTag[362], methyl-ATAC-seq[363], ATAC-Me[364]) and chromatin accessibility profiling with both DNA methylation and transcriptome measurements (scNMT-seq[365]).

Several technical challenges have so far limited the widespread application of multi-omic methods. Sample fixation, reaction conditions and other experimental parameters are often not compatible for multiple omic assays, complicating the optimization of joint protocols. Given that single-cell omic methods often suffer from sensitivity issues — meaning a low number of detected features (such as genes or regions) per single cell — running and combining two such methods could result in a very small set of overlapping features. Profiling multiple molecular layers raises the non-trivial computational challenge of integrating the data sets. Methods that can handle the harmonization of bulk and single-cell multi-omic measurements have recently been developed (MOFA[366], *Seurat v3* (REF.[240])). A key feature required for future computational methods is flexibility; methods need to handle data sets 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, such as putative enhancers, and gene transcription from the same cell, for example, would facilitate the linking of enhancers to their potential target genes. We anticipate important developments in both experimental and computational multi-omics approaches in the coming years.

The function of accessible chromatin regions can also be probed by perturbation, for example by mutating key TFs. Single-cell accessibility profiling can detect 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

Morphogen gradients
Gradients of signalling molecules within developing tissues and embryos, which illicit different responses across the gradient, leading to diverse outcomes in terms of cell fate decisions, controlling pattern formation during embryogenesis.

performed in cell culture models by coupling CRISPR screening with scATAC-seq (Perturb-ATAC[367]). In more complex systems, where high-throughput targeted mutagenesis is not feasible, natural sequence variation can be exploited for large-scale perturbation. In this context, profiling accessibility within and between species provides insights into regulatory variation and functionality. These approaches, in combination with multi-omics measurements, may lead to more accurate and predictive models of gene expression, and to a causal understanding of enhancer and TF function.

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[84] mitigates this problem by performing the Tn5 reaction in situ on microscope slides and using fluorescent adaptors that are compatible with both imaging and sequencing. sciMAP-ATAC[368] provides medium-level spatial mapping of single-cell chromatin accessibility profiles for a tissue by taking a select amount of microbiopsies of a tissue prior to the sciATAC-seq workflow. Further integration of ATAC-seq with high-throughput fluorescence in situ hybridization and other imaging-based methods will lead to new ways of interrogating the genome of complex systems in situ after stimuli and perturbations[368,369]. Such developments hold promise to advance discovery in multiple fields. For example, in the context of developmental biology, it would help to decode the functional impact of morphogen gradients and cellular signal transduction by measuring the regulatory response in the cells receiving the signal while maintaining information on each cell's spatial positioning to both the source signal and their neighbouring cells. Integration with multi-omics measurements will lead to the generation of virtual models of developing embryos with enhanced resolution and predictive power. In the medical setting, this could reveal the relationship between cell growth and spatial positions within a tumour, the dependencies between the point of injury or infection, and the efficacy of drugs to elicit cellular responses depending on the cells' position, to name but a few. Given the sensitivity of these methods and the rapid speed with which they are being developed, this will open up new, exciting avenues for diagnosis, prognosis and therapeutic intervention.

Published online: 21 January 2021

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).
This paper represents an extensive map of DHSs, identifying and annotating nearly 3 million DHSs, and thereby demonstrates relationships between chromatin accessibility, transcription and TF binding patterns.
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. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).
14. Cook, P. R. The organization of replication and transcription. *Science* **284**, 1790–1795 (1999).
15. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **18**, 407–422 (2017).
16. Gillette, T. G. & Hill, J. A. Readers, writers, and erasers: chromatin as the whiteboard of heart disease. *Circ. Res.* **116**, 1245–1253 (2015).
17. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463**, 474–484 (2010).
18. Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11**, 1587–1598 (2003).
19. Reinke, H. & Hörz, W. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol. Cell* **11**, 1599–1607 (2003).
20. 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).
21. 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).
22. 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).
23. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
This review article describes the main properties of pioneer factors and their important role in establishing the chromatin landscape and in enabling cellular reprogramming.
24. Zhu, F. et al. The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
25. Hendrich, B. & Bickmore, W. Human diseases with underlying defects in chromatin structure and modification. *Hum. Mol. Genet.* **10**, 2233–2242 (2001).
26. Matsumoto, L. et al. CpG demethylation enhances α-synuclein expression and affects the pathogenesis of Parkinson's disease. *PLoS ONE* **5**, e15522 (2010).
27. Schwartzentruber, J. et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).
28. Vinagre, J. et al. Frequency of TERT promoter mutations in human cancers. *Nat. Commun.* **4**, 2185 (2013).
29. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
30. Davis, C. A. et al. The Encyclopedia of DNA Elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
31. Stunnenberg, H. G. et al. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell* **167**, 1145–1149 (2016).
32. Bernstein, B. E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
33. Adams, D. et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30**, 224–226 (2012).
34. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
35. Boyle, A. P. et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
This study is the first to apply genome-wide sequencing to profile chromatin accessibility, by means of DNase-seq.
36. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
37. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
38. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, (2017).
39. 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).
40. Hesselberth, J. R. et al. Global mapping of protein–DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
41. Kelly, T. K. et al. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).
This study develops the first genome-wide assay for single-molecule chromatin accessibility profiling, relying on methyltransferase enzymes that preferentially modify accessible DNA.
42. Schones, D. E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
43. Taberlay, P. C. et al. Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell* **147**, 1283–1294 (2011).

44. 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).
45. Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).
This pioneering work in the field of regulatory genomics shows that genomic regions of active transcription are particularly sensitive to digestion by DNase I, indicating a more permissive form of the chromatin.
46. 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).
47. 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).
This article establishes DNase footprinting as a method to study the sequence-specific binding of proteins to DNA.
48. Kemper, B., Jackson, P. D. & Felsenfeld, G. Protein-binding sites within the 5′ DNase I-hypersensitive region of the chicken α d-globin gene. *Mol. Cell. Biol.* **7**, 2059–2069 (1987).
49. Vierstra, J. et al. Global reference mapping of human transcription factor footprints. *Nature* **583**, 729–736 (2020).
50. 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).
51. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
52. West, J. A. et al. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat. Commun.* **5**, 4719 (2014).
This work shows that changes in nucleosome occupancy during cellular differentiation are enriched at regulatory regions and co-localize with binding sites of key developmental regulators.
53. Reddington, J. et al. Lineage resolved enhancer and promoter usage during a time-course of embryogenesis. *Dev. Cell* **55**, 648–664 (2020).
54. Al-Ali, R. et al. Single-nucleus chromatin accessibility reveals intratumoral epigenetic heterogeneity in IDH1 mutant gliomas. *Acta Neuropathol. Commun.* **7**, 201 (2019).
55. Cusanovich, D. A. et al. The *cis*-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018).
56. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324.e18 (2018).
57. Fullard, J. F. et al. An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28**, 1243–1252 (2018).
58. Jin, W. et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* **528**, 142–146 (2015).
59. 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).
60. Pijuan-Sala, B. et al. Single-cell chromatin accessibility maps reveal regulatory programs driving early mouse organogenesis. *Nat. Cell Biol.* **22**, 487–497 (2020).
61. Preissl, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).
62. 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).
63. Wang, X. et al. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat. Commun.* **9**, 5380 (2018).
64. 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).
65. Graybuck, L. T. et al. Enhancer viruses and a transgenic platform for combinatorial cell subclass-specific labeling. Preprint at *bioRxiv* https://doi.org/10.1101/525014 (2019).
66. Hafez, D. et al. McEnhancer: predicting gene expression via semi-supervised assignment of enhancers to target genes. *Genome Biol.* **18**, 199 (2017).
67. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
68. 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).
69. Ricci, W. A. et al. Widespread long-range *cis*-regulatory elements in the maize genome. *Nat. Plants* **5**, 1237–1249 (2019).
70. 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).
71. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
72. 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).
73. Lu, F. et al. Establishing chromatin regulatory landscape during mouse preimplantation development. *Cell* **165**, 1375–1388 (2016).
74. 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).
75. Lazarovici, A. et al. Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl Acad. Sci. USA* **110**, 6376–6381 (2013).
76. Suck, D., Lahm, A. & Oefner, C. Structure refined to 2A of a nicked DNA octanucleotide complex with DNase I. *Nature* **332**, 464–468 (1988).
77. 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).
78. Sung, M.-H., Baek, S. & Hager, G. L. Genome-wide footprinting: ready for prime time? *Nat. Methods* **13**, 222–228 (2016).
79. 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).
80. Goryshin, I. Y. & Reznikoff, W. S. Tn5 in vitro transposition. *J. Biol. Chem.* **273**, 7367–7374 (1998).
81. 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).
82. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
83. Wu, J. et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).
84. Chen, X. et al. ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
85. 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).
86. 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).
87. Sato, S. et al. Biochemical analysis of nucleosome targeting by Tn5 transposase. *Open Biol.* **9**, 190116 (2019).
88. Karabacak Calviello, A., Hirsekorn, A., Wurmus, R., Yusuf, D. & Ohler, U. Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol.* **20**, 42 (2019).
89. 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).
90. Montefiori, L. et al. Reducing mitochondrial reads in ATAC-seq using CRISPR/Cas9. *Sci. Rep.* **7**, 2451 (2017).
91. Sos, B. C. et al. Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. *Genome Biol.* **17**, 20 (2016).
92. Chereji, R. V., Bryson, T. D. & Henikoff, S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol.* **20**, 198 (2019).
93. Chang, P., Gohain, M., Yen, M.-R. & Chen, P.-Y. Computational methods for assessing chromatin hierarchy. *Comput. Struct. Biotechnol. J.* **16**, 43–53 (2018).
94. Kensche, P. R. et al. The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Res.* **44**, 2110–2124 (2016).
95. 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).
96. Lai, B. et al. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* **562**, 281–285 (2018).
97. 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. USA* **100**, 7743–7748 (2003).
98. Jessen, W. J. et al. Mapping chromatin structure in vivo using DNA methyltransferases. *Methods* **33**, 68–80 (2004).
99. 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).
100. 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).
101. 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.* **1288**, 123–141 (2015).
102. 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).
103. Shipony, Z. et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).
104. 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).
105. Sönmezer, C. et al. Molecular co-occupancy identifies transcription factor binding cooperativity in iivo. *Mol. Cell* **20**, S1097–2765 (2020).
106. Yang, Y. et al. Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). *BMC Genomics* **16**, 350 (2015).
107. Payne, A., Holmes, N., Rakyan, V. & Loose, M. BulkVis: a graphical viewer for Oxford Nanopore bulk FAST5 files. *Bioinformatics* **35**, 2193–2198 (2019).
108. Tyler, A. D. et al. Evaluation of Oxford Nanopore's minion sequencing device for microbial whole genome sequencing applications. *Sci. Rep.* **8**, 10931 (2018).
109. Mahmoud, M., Zywicki, M., Twardowski, T. & Karlowski, W. M. Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. *Genomics* **111**, 43–49 (2019).
110. Pfeiffer, F. et al. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* **8**, 10950 (2018).
111. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 (2020).
112. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics Proteom. Bioinforma.* **13**, 278–289 (2015).
113. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
114. Wang, Y. et al. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* **29**, 1329–1342 (2019).
115. Liu, Y. et al. Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS. *Genome Biol.* **21**, 54 (2020).
116. 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).
117. Abdulhay, N. J. et al. Massively multiplex single-molecule oligonucleosome footprinting. *eLife* **9**, e59404 (2020).
118. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).

119. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
120. Robertson, G. et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
121. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).
122. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
123. 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).
124. 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).
125. Lara-Astiaso, D. et al. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
126. Bonn, S. et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012).
127. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
128. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
129. Kuo, M. H. & Allis, C. D. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* **19**, 425–433 (1999).
130. O'Neill, L. P. & Turner, B. M. Immunoprecipitation of native chromatin: NChIP. *Methods* **31**, 76–82 (2003).
131. Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **25**, 99–104 (2000).
132. 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).
133. Dahl, J. A. et al. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548–552 (2016).
134. Ng, J.-H. et al. In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev. Cell* **24**, 324–333 (2013).
135. Zhang, B. et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 (2016).
136. 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).
137. 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).
138. Harada, A. et al. A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.* **21**, 287–296 (2019).
139. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
140. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
141. Ku, W. L. et al. Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat. Methods* **16**, 323–325 (2019).
142. Wang, Q. et al. CoBATCH for high-throughput single-cell epigenomic profiling. *Mol. Cell* **76**, 206–216.e7 (2019).
143. Hainer, S. J., Boskovic, A., Rando, O. J. & Fazzio, T. G. Profiling of pluripotency factors in individual stem cells and early embryos. *Cell* https://doi.org/10.1101/286351 (2018).
144. Rotem, A. et al. Single-cell ChIP–seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
145. Grosselin, K. et al. High-throughput single-cell ChIP–seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).
146. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015). **This study is one of the first to perform genome-wide chromatin accessibility profiling at a single-cell level, thus spearheading the now rising use of scATAC-seq**.
147. 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).
148. Cusanovich, D. A. et al. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015). **This study is one of the first to perform genome-wide chromatin accessibility profiling at a single-cell level, thus spearheading the now rising use of scATAC-seq**.
149. Lareau, C. A. et al. Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **37**, 916–924 (2019).
150. 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).
151. Mezger, A. et al. High-throughput chromatin accessibility profiling at single-cell resolution. *Nat. Commun.* **9**, 3647 (2018).
152. Domcke, S. et al. A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020).
153. Ma, S. et al. Chromatin potential identified by shared single cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 (2020).
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* https://doi.org/10.5281/ZENODO.156534 (2016).
157. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
158. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17**, 10 (2011).
159. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
160. Aronesty, E. ea-utils, FASTQ processing utilities https://expressionanalysis.github.io/ea-utils/ (2011).
161. Pass, D. A. et al. Genome-wide chromatin mapping with size resolution reveals a dynamic sub-nucleosomal landscape in *Arabidopsis*. *PLOS Genet.* **13**, e1006988 (2017).
162. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
163. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
164. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).
165. Ou, J. et al. ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics* **19**, 169 (2018).
166. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
167. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
168. Yates, A. D. et al. Ensembl 2020. *Nucleic Acids Res.* **48**, 682–688 (2019).
169. Buels, R. et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* **17**, 66 (2016).
170. Hofmeister, B. T. & Schmitz, R. J. Enhanced JBrowse plugins for epigenomics data visualization. *BMC Bioinforma.* **19**, 159 (2018).
171. Zhang, Y. et al. Model-based analysis of ChIP–seq (MACS). *Genome Biol.* **9**, R137 (2008).
172. 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).
173. Tarbell, E. D. & Liu, T. HMMRATAC: a Hidden Markov ModeleR for ATAC-seq. *Nucleic Acids Res.* **47**, e91–e91 (2019).
174. Gaspar, J. M. Genrich: detecting sites of genomic enrichment. *Github* https://github.com/jsh58/Genrich (2018).
175. Boyle, A. P., Guinney, J., Crawford, G. E. & Furey, T. S. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatic.* **24**, 2537–2538 (2008).
176. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264–268 (2011).
177. Koohy, H., Down, T. A., Spivakov, M. & Hubbard, T. A comparison of peak callers used for DNase-seq data. *PLoS ONE* **9**, e96303 (2014).
178. Boleu, N., Kundaje, A. & Bickel, P. J. Irreproducible Discovery Rate (IDR). https://github.com/nboley/idr (2016).
179. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
180. 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**, 517–532 (2015).
181. The ENCODE Project Consortium. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020). **This paper summarizes years of effort from the ENCODE project, which have led to a recourse of almost 1 million human and more than 300,000 mouse candidate regulatory elements**.
182. 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).
183. 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).
184. Stark, R. & Brown, G. Differential binding analysis of ChIP- Seq peak data. https://bioconductor.org/packages/release/bioc/html/DiffBind.html (2020).
185. 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).
186. Liang, K. & Keles, S. Detecting differential binding of transcription factors with ChIP–seq. *Bioinformatics* **28**, 121–122 (2012).
187. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
188. Reske, J. J., Wilson, M. R. & Chandler, R. L. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics Chromatin* **13**, 22 (2020).
189. Lun, A. T. L. csaw: a Bioconductor package for differential binding analysis of ChIP–seq data using sliding windows. *Nucleic Acids Res.* **44**, e45 (2016).
190. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
191. Bravo González-Blas, C. et al. cisTopic: *cis*-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400 (2019).
192. Gandolfi, F. & Tramontano, A. A computational approach for the functional classification of the epigenome. *Epigenetics Chromatin* **10**, 26 (2017).
193. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
194. 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).
195. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP–seq and ChIP-chip data. *BMC Bioinforma.* **11**, 237 (2010).
196. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinforma.* **14**, 128 (2013).
197. 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**, e114 (2012).
198. 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).
199. Layer, R. M. et al. GIGGLE: a search engine for large-scale integrated genome analysis. *Nat. Methods* **15**, 123–126 (2018).
200. Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587–589 (2016).

201. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
202. 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).
203. Hoffman, M. M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat. Methods* **9**, 473–476 (2012).
204. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
205. 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). **This article presents the first use of a position weight matrix, the currently most widely used model for representing binding sites of a TF**.
206. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, 87–92 (2019).
207. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
208. 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).
209. 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).
210. Thomas-Chollier, M. et al. RSAT peak-motifs: motif analysis in full-size ChIP–seq datasets. *Nucleic Acids Res.* **40**, e31–e31 (2012).
211. 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).
212. Kelley, D. R., Snoek, J. & Rinn, J. L. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res.* **26**, 990–999 (2016).
213. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **12**, 931–934 (2015).
214. Shrikumar, A., Greenside, P. & Kundaje, A. Learning important features through propagating activation differences. Preprint at *ArXiv* https://arxiv.org/abs/1704.02685 (2017).
215. Minnoye, L. et al. Cross-species analysis of enhancer logic using deep learning. *Genome Res.* **30**, 1815–1834 (2020).
216. Baek, S. & Sung, M.-H. in *Statistical Genomics* Vol. 1418 (eds Mathé, E. & Davis, S.) 225–240 (Springer, 2016).
217. Neph, S. et al. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
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. 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).
226. Li, Z. et al. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).
227. Quach, B. & Furey, T. S. DeFCoM: analysis and modeling of transcription factor binding sites using a motif-centric genomic footprinter. *Bioinformatics* **33**, 956–963 (2017).
228. Bentsen, M. et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).
229. Chen, K. et al. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351 (2013).
230. 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).
231. Schep, A. N. et al. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res.* **25**, 1757–1770 (2015).
232. Zhong, J. et al. Mapping nucleosome positions using DNase-seq. *Genome Res.* **26**, 351–364 (2016).
233. Baker, S. M., Rogerson, C., Hayes, A. & Sharrocks, A. D. Classifying cells with Scasat — a tool to analyse single-cell ATAC-seq. *Nucleic Acids Res.* **47**, e10 (2017).
234. de Boer, C. G. & Regev, A. BROCKMAN: deciphering variance in epigenomic regulators by *k*-mer factorization. *BMC Bioinforma*. **19**, (2018).
235. Fang, R. et al. Fast and accurate clustering of single cell epigenomes reveals *cis*-regulatory elements in rare cell types. Preprint at *bioRxiv* https://doi.org/10.1101/615179 (2019).
236. Granja, J. M. et al. ArchR: an integrative and scalable software package for single-cell chromatin accessibility analysis. Preprint at *bioRxiv* https://doi.org/10.1101/2020.04.28.066498 (2020).
237. Ji, Z., Zhou, W. & Ji, H. Single-cell regulome data analysis by SCRAT. *Bioinformatics* **33**, 2930–2932 (2017).
238. Pliner, H. A. et al. Cicero predicts *cis*-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871.e8 (2018).
239. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
240. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
241. Zamanighomi, M. et al. Unsupervised clustering and epigenetic classification of single cells. *Nat. Commun.* **9**, 2410 (2018).
242. Wang, C. et al. Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol.* **21**, 198 (2020).
243. Fang, R. et al. SnapATAC: a comprehensive analysis package for single cell ATAC-seq. Preprint at *bioRxiv* https://doi.org/10.1101/615179 (2019).
244. Chen, H. et al. Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biol.* **20**, 241 (2019).
245. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
246. Danese, A., Richter, M. L., Fischer, D. S., Theis, F. J. & Colomé-Tatché, M. EpiScanpy: integrated single-cell epigenomic analysis. Preprint at *bioRxiv* https://doi.org/10.1101/648097 (2019).
247. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
248. Baek, S. & Lee, I. Single-cell ATAC sequencing analysis: from data preprocessing to hypothesis generation. *Comput. Struct. Biotechnol. J.* **18**, 1429–1439 (2020).
249. Polański, K. et al. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2019).
250. Hie, B., Bryson, B. & Berger, B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).
251. 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).
252. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
253. Luecken, M. et al. Benchmarking atlas-level data integration in single-cell genomics. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.22.111161 (2020).
254. 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).
255. Chen, H. et al. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat. Commun.* **10**, 1903 (2019).
256. Trapnell, C. & Cacchiarelli, D. Monocle. https://github.com/cole-trapnell-lab/monocle-release (2020).
257. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018). **This work establishes the first scalable genome-wide technique that allows simulations profiling transcription and chromatin accessibility on a single-cell level, illustrating the advantage of a multi-omics single-cell assay to link regulatory elements to regulated genes**.
258. Kaplan, N. et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
259. Segal, E. et al. A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
260. Gaffney, D. J. et al. Controls of nucleosome positioning in the human genome. *PLoS Genet.* **8**, e1003036 (2012).
261. Tillo, D. et al. High nucleosome occupancy is encoded at human regulatory sequences. *PLoS ONE* **5**, e9129 (2010).
262. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet.* **21**, 292–310 (2020).
263. 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).
264. 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).
265. Sun, Y. et al. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Res.* **25**, 1703–1714 (2015).
266. Gao, L. et al. Chromatin accessibility landscape in human early embryos and its association with evolution. *Cell* **173**, 248–259 (2018).
267. Lee, M. T. et al. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).
268. Leichsenring, M., Maes, J., Mössner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005–1009 (2013).
269. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **293**, 13795–13804 (2018).
270. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244 (2018).
271. 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).
272. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).
273. Buenrostro, J. D. et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell* **173**, 1535–1548.e16 (2018).
274. Yoshida, H. et al. The *cis*-regulatory atlas of the mouse immune system. *Cell* **176**, 897–912.e20 (2019).
275. Gosselin, D. et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
276. Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
277. Satpathy, A. T. et al. Transcript-indexed ATAC-seq for precision immune profiling. *Nat. Med.* **24**, 580–590 (2018).
278. 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).
279. Koues, O. I. et al. Distinct gene regulatory pathways for human innate versus adaptive lymphoid cells. *Cell* **165**, 1134–1146 (2016).

280. Shih, H.-Y. et al. Developmental acquisition of regulomes underlies innate lymphoid cell functionality. *Cell* **165**, 1120–1133 (2016).
281. Youngblood, B. et al. Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature* **552**, 404–409 (2017).
282. van der Veeken, J. et al. Memory of inflammation in regulatory T cells. *Cell* **166**, 977–990 (2016).
283. Netea, M. G. et al. Defining trained immunity and its role in health and disease. *Nat. Rev. Immunol.* **20**, 375–388 (2020).
284. Novakovic, B. et al. β-Glucan reverses the epigenetic state of LPS-induced immunological tolerance. *Cell* **167**, 1354–1368.e14 (2016).
285. Saeed, S. et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* **345**, 1251086 (2014).
286. Krausgruber, T. et al. Structural cells are key regulators of organ-specific immune responses. *Nature* **583**, 296–302 (2020).
287. de la Torre-Ubieta, L. et al. The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell* **172**, 289–304.e18 (2018).
288. Prescott, S. L. et al. Enhancer divergence and *cis*-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).
289. Trevino, A. E. et al. Chromatin accessibility dynamics in a model of human forebrain development. *Science* **367**, eaay1645 (2020).
290. Nott, A. et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
291. Yin, S. et al. Transcriptomic and open chromatin atlas of high-resolution anatomical regions in the rhesus macaque brain. *Nat. Commun.* **11**, 474 (2020).
292. 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).
293. 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).
294. 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).
295. 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).
296. Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
297. Beekman, R. et al. The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. *Nat. Med.* **24**, 868–880 (2018).
298. Ott, C. J. et al. Enhancer architecture and essential core regulatory circuitry of chronic lymphocytic leukemia. *Cancer Cell* **34**, 982–995.e7 (2018).
299. Rendeiro, A. F. et al. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat. Commun.* **7**, 11938 (2016).
300. 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).
301. 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).
302. Schmidl, C. et al. Combined chemosensitivity and chromatin profiling prioritizes drug combinations in CLL. *Nat. Chem. Biol.* **15**, 232–240 (2019).
303. Akhtar-Zaidi, B. et al. Epigenomic enhancer profiling defines a signature of colon cancer. *Science* **336**, 736–739 (2012).
304. Cohen, A. J. et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat. Commun.* **8**, 14400 (2017).
305. Guilhamon, P. et al. Single-cell chromatin accessibility in glioblastoma delineates cancer stem cell heterogeneity predictive of survival. Preprint at *bioRxiv* https://doi.org/10.1101/370726 (2018).
306. Tome-Garcia, J. et al. Analysis of chromatin accessibility uncovers TEAD1 as a regulator of migration in human glioblastoma. *Nat. Commun.* **9**, 4020 (2018).
307. Ooi, W. F. et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat. Commun.* **7**, 12983 (2016).
308. Denny, S. K. et al. Nfib promotes metastasis through a widespread increase in chromatin accessibility. *Cell* **166**, 328–342 (2016).
309. Wang, Z. et al. The open chromatin landscape of non-small cell lung carcinoma. *Cancer Res.* **79**, 4840–4854 (2019).
310. 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).
311. 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).
312. Torchia, J. et al. Integrated (epi)-genomic analyses identify subgroup-specific therapeutic targets in CNS RHABDOID tumors. *Cancer Cell* **30**, 891–908 (2016).
313. Halbritter, F. et al. Epigenomics and single-cell sequencing define a developmental hierarchy in langerhans cell histiocytosis. *Cancer Discov.* **9**, 1406–1421 (2019).
314. 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).
315. Sen, D. R. et al. The epigenetic landscape of T cell exhaustion. *Science* **354**, 1165–1169 (2016).
316. Ghoneim, H. E. et al. De novo epigenetic programs inhibit PD-1 blockade-mediated T cell rejuvenation. *Cell* **170**, 142–157.e19 (2017).
317. Pauken, K. E. et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).
318. 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.* **20**, 651–668 (2020).
319. Boyd, M. et al. Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies. *Nat. Commun.* **9**, 1661 (2018).
320. Ai, R. et al. Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes. *Nat. Commun.* **9**, 1921 (2018).
321. 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).
322. Bryois, J. et al. Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, 3121 (2018).
323. Sun, W. et al. Histone acetylome-wide association study of autism spectrum disorder. *Cell* **167**, 1385–1397.e11 (2016).
324. 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).
325. 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).
326. Degner, J. F. et al. DNaseI sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012).
327. Gate, R. E. et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* **50**, 1140–1150 (2018).
328. Jacobs, J. et al. The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat. Genet.* **50**, 1011–1020 (2018).
329. Atak, Z. K. et al. Prioritization of enhancer mutations by combining allele-specific chromatin accessibility with deep learning. Preprint at *bioRxiv* https://doi.org/10.1101/2019.12.21.885806 (2019).
330. Roscito, J. G. et al. Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. *Nat. Commun.* **9**, 4737 (2018).
331. 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).
332. Stone, J. R. & Wray, G. A. Rapid evolution of *cis*-regulatory sequences via local point mutations. *Mol. Biol. Evol.* **18**, 1764–1770 (2001).
333. Lu, Z. et al. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* **5**, 1250–1259 (2019).
334. 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).
335. Sebé-Pedrós, A. et al. The dynamic regulatory genome of capsaspora and the origin of animal multicellularity. *Cell* **165**, 1224–1237 (2016).
336. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
337. Kolesnikov, N. et al. ArrayExpress update — simplifying data submissions. *Nucleic Acids Res.* **43**, D1113–D1116 (2015).
338. Leinonen, R., Sugawara, H., Shumway, M. & International Nucleotide Sequence Database Collaboration. The Sequence Read Archive. *Nucleic Acids Res.* **39**, D19–D21 (2011).
339. Leinonen, R. et al. The European Nucleotide Archive. *Nucleic Acids Res.* **39**, D28–D31 (2011).
340. Kaminuma, E. et al. DDBJ launches a new archive database with analytical tools for next-generation sequence data. *Nucleic Acids Res.* **38**, D33–D38 (2010).
341. Lappalainen, I. et al. The European Genome–phenome Archive of human data consented for biomedical research. *Nat. Genet.* **47**, 692–695 (2015).
342. Mailman, M. D. et al. The NCBI dbGaP database of Genotypes and Phenotypes. *Nat. Genet.* **39**, 1181–1186 (2007).
343. Bujold, D. et al. The International Human Epigenome Consortium data portal. *Cell Syst.* **3**, 496–499.e2 (2016).
344. Oki, S. et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP–seq data. *EMBO Rep.* **19**, e46255 (2018).
345. Chèneby, J., Gheorghe, M., Artufel, M., Mathelier, A. & Ballester, B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP–seq experiments. *Nucleic Acids Res.* **46**, D267–D275 (2018).
346. Raney, B. J. et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC genome browser. *Bioinformatics* **30**, 1003–1005 (2014).
347. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
348. Davie, K. et al. A single-cell transcriptome atlas of the aging *Dosophila* brain. *Cell* **174**, 982–998.e20 (2018).
349. 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).
350. Papatheodorou, I. et al. Expression Atlas update: from tissues to single cells. *Nucleic Acids Res.* **48**, 77–83 (2019).
351. Avsec, Ž. et al. The Kipoi repository accelerates community exchange and reuse of predictive models for genomics. *Nat. Biotechnol.* **37**, 592–600 (2019).
352. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
353. 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).
354. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394 (2011).
355. 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).
356. Colli, M. L. et al. An integrated multi-omics approach identifies the landscape of interferon-α-mediated responses of human pancreatic β cells. *Nat. Commun.* **11**, 2584 (2020).
357. 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).
358. Maheshwari, S. et al. Massively parallel simultaneous profiling of the transcriptome and epigenomic landscape at single cell resolution. *10x Genomics* https://pages.10xgenomics.com/rs/446-PBO-704/images/10x_AGBT_Poster_2020_Massively-parallel-simultaneous-profiling-of-the-transcriptomic-and-epigenomic-landscape-at-single-cell-resolution.pdf (2020).
359. Chen, X. et al. Joint single-cell DNA accessibility and protein epitope profiling reveals environmental regulation of epigenomic heterogeneity. *Nat. Commun.* **9**, 4590 (2018).

360. Pott, S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* **6**, e23203 (2017).
361. Guo, F. et al. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res.* **27**, 967–988 (2017).
362. Lhoumaud, P. et al. EpiMethylTag: simultaneous detection of ATAC-seq or ChIP–seq signals with DNA methylation. *Genome Biol.* **20**, 248 (2019).
363. 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).
364. 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).
365. 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).
366. Argelaguet, R. et al. Multi-Omics Factor Analysis — a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, e8124 (2018).
367. Rubin, A. J. et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* **176**, 361–376.e17 (2019).
368. Thornton, C. A. et al. Spatially-mapped single-cell chromatin accessibility. Preprint at *bioRxiv* https://doi.org/10.1101/815720 (2019).
369. Packer, J. & Trapnell, C. Single-cell multi-omics: an engine for new quantitative models of gene regulation. *Trends Genet.* **34**, 653–665 (2018).
370. Ponnaluri, V. K. C. et al. NicE-seq: high resolution open chromatin profiling. *Genome Biol.* **18**, 122 (2017).
371. Giresi, P. G. & Lieb, J. D. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (formaldehyde assisted isolation of regulatory elements). *Methods* **48**, 233–239 (2009).
372. Lai, B. et al. TrAC-looping measures genome structure and chromatin accessibility. *Nat. Methods* **15**, 741–747 (2018).
373. Spracklin, G. & Pradhan, S. Protect-seq: genome-wide profiling of nuclease inaccessible domains reveals physical properties of chromatin. *Nucleic Acids Res.* **48**, e16 (2020).
374. Tchasovnikarova, I. A. et al. Hyperactivation of HUSH complex function by Charcot–Marie–Tooth disease mutation in MORC2. *Nat. Genet.* **49**, 1035–1044 (2017).
375. 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).
376. 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**, e32341 (2018).
377. 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).
378. Ishii, H., Kadonaga, J. T. & Ren, B. MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proc. Natl Acad. Sci. USA* **112**, E3457–E3465 (2015).
379. 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).
380. 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).
381. 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).
382. Oberbeckmann, E. et al. Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.* **29**, 1996–2009 (2019).
383. 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).
384. Voong, L. N. et al. Insights into nucleosome organization in mouse embryonic stem cells through chemical mapping. *Cell* **167**, 1555–1570.e15 (2016).
385. 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).
386. 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. USA* **93**, 1370–1375 (1996).
387. Pott, S. & Lieb, J. D. Single-cell ATAC-seq: strength in numbers. *Genome Biol.* **16**, 172–172 (2015).

Acknowledgements
This work was supported by the Division of Intramural Research, National Heart, Lung and Blood Institute, National Institutes of Health (NIH) (K.Z.) and 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 (DBI-1905869), C.B. by a European Research Council (ERC) Starting Grant (European Union's Horizon 2020 research and innovation programme, grant agreement no. 679146), E.E.M.F. by an ERC Advanced grant (DeCRypT) and S.A. by an ERC Consolidator Grant (cis_CONTROL). W.J.G. acknowledges support as a Chan-Zuckerberg Biohub Investigator, and from NIH grants UM1-HG009436, P50-HG007735, UM1-HG009442 and U19-AI057266. The authors thank J. Demeulemeester for insights into the long-read sequencing platforms.

Author contributions
Introduction (S.A., L.M.); Experimentation (G.K.M., L.P., A.P.M., W.J.G., K.Z., R.J.S.); Results (L.M., S.A., G.K.M., W.J.G.); Applications (G.K.M., T.K., A.P.M., R.J.S., C.B., W.J.G.); Reproducibility and data deposition (A.P.M., R.J.S.); Limitations and optimizations (G.K.M., W.J.G.); Outlook (S.S., E.E.M.F.); Overview of the Primer (S.A., L.M.).

Competing interests
C.B. is an inventor on a patent describing the ChIPmentation assay. R.J.S. is a co-founder of REquest Genomics, LLC, a company that provides epigenomics services. W.J.G is an inventor on a patent describing Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), a consultant for 10x Genomics and Guardant Health, and a co-founder of Protillion biosciences. All other authors declare no competing interests.

Peer review information
Nature Reviews Methods Primers thanks T. Liu, B. Treutlein, L. Zhu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s43586-020-00008-9.

RELATED LINKS
ENCODE ATAC-seq Data Standards and Processing Pipeline: https://www.encodeproject.org/atac-seq/
FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/