APPLICATIONS OF NEXT-GENERATION SEQUENCING

High-resolution digital profiling of the epigenome

Gabriel E. Zentner[1] and Steven Henikoff[1,2]

Abstract | The widespread adoption of short-read DNA sequencing as a digital epigenomic readout platform has motivated the development of genome-wide tools that achieve base-pair resolution. New methods for footprinting and affinity purification of nucleosomes, RNA polymerases, chromatin remodellers and transcription factors have increased the resolution of epigenomic profiling by two orders of magnitude, leading to new insights into how the chromatin landscape affects gene regulation. These digital epigenomic tools have also been applied to directly profile both turnover kinetics and transcription *in situ*. In this Review, we describe how these new genome-wide tools allow interrogation of diverse aspects of the epigenome.

The epigenome is a complex layer of regulatory information superimposed on the genome and includes nucleosome occupancy[1], positioning[1], composition[2], modification[3] and dynamics[4], as well as DNA methylation[5]; it is shaped by the action of transcription factors (TFs)[6], ATP-dependent chromatin remodellers[7], non-coding RNAs[8] and other factors. There has been great interest in generating detailed epigenomic maps to understand both fundamental cellular processes and the molecular bases of human disease, and this is exemplified by the large-scale epigenomic mapping efforts of consortia such as Encyclopedia of DNA Elements (ENCODE)[9] and the Epigenomics Roadmap[10]. Beyond these efforts, epigenomic techniques have become commonplace in many areas of biological research. For example, a search of the Sequence Read Archive (SRA) for ChIP–seq (chromatin immunoprecipitation followed by high-throughput sequencing) yields >24,000 data sets.

Massively parallel sequencing is revolutionizing genomics, although the short length and low accuracy of sequence reads relative to Sanger sequencing have been limiting for identifying polymorphisms and mutations[11]. However, short-read sequencing in the 25–75 bp range is ideally suited for most epigenomic applications, in which only enough accuracy is needed to confidently map a DNA fragment to its correct position in a reference genome. To the extent that a sequence read precisely corresponds to the location of an epigenomic feature, base-pair resolution delineation of epigenomic features is feasible.

The potential for using short-read sequencing for epigenomics is enhanced by the precipitous decrease in sequencing costs that has occurred over the past several years. For instance, when the first ChIP–seq studies[12–14] were published in 2007, sequencing costs were on the order of ~US$800 per megabase (see DNA sequencing costs). By the end of 2013, a megabase of sequencing costs less than $0.1. Recently, this decrease in the cost of sequencing has spurred the development of a plethora of genome-wide techniques that use short-read sequencing as a readout.

DNA methylation is an epigenomic component that is routinely mapped at base-pair resolution with bisulphite sequencing[15]. However, the resolution of most epigenomic techniques, particularly those used to map the occupancy and dynamics of nucleosomes and non-histone factors, has been limited by the methods used to prepare chromatin. Why is base-pair resolution important in such cases? For analysis of TF binding, base-pair resolution is essential for determining the precise DNA sequence bound by a TF and for identifying single binding sites within a cluster of closely spaced sites. Precise mapping of nucleosome positions is also important for understanding every DNA-templated process, as nucleosomes occlude TF binding sites and other regulatory sequences, and single-base shifts in nucleosome positioning can alter chromatin structure[16]. In this Review, we discuss how recently developed epigenomic methods can be used to probe various aspects of the epigenome at or near base-pair resolution, with a focus on nucleosome and non-histone factor occupancy and dynamics.

[1]Basic Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N, Seattle, Washington 98109, USA. [2]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N, Seattle, Washington 98109, USA. Correspondence to S.H. e-mail: steveh@fhcrc.org

Footprinting

One of the first properties of chromatin to be assayed was the accessibility of specific loci, as defined by their sensitivity to cleavage by nonspecific nucleases. Two of the most widely used such nucleases are micrococcal nuclease (MNase)[17] and deoxyribonuclease I (DNase I)[18]. At the start of the epigenomics era, MNase and DNase I cleavage was adapted to microarray readout platforms[19–21]. In parallel, sonication-based methods to assay chromatin accessibility were developed. These methods, including formaldehyde-assisted isolation of regulatory elements (FAIRE)[22] and sonication of crosslinked chromatin sequencing (Sono-seq)[23], rely on the differential sensitivity of 'open' (nucleosome-depleted) and 'closed' (nucleosome-occupied) chromatin to breakage by sonication. Given the limited resolution of microarray analysis, both cleavage- and sonication-based methods gave similar resolution for measuring chromatin accessibility.

The use of enzymatic methods for assaying chromatin accessibility has an additional benefit: nucleosomes and other chromatin-associated proteins impede enzymatic cleavage or modification of DNA, creating 'footprints' that correspond to protein-bound DNA sequences with base-pair resolution. By contrast, the resolution of sonication-based methods for assaying chromatin accessibility is intrinsically limited by the heterogeneous range of fragment sizes generated by sonication, and these techniques thus measure only chromatin accessibility without footprinting bound factors. With the development of high-throughput sequencing, there has been renewed interest in adapting cleavage-based methods of chromatin profiling for digital epigenomic analysis. We discuss below four such techniques and one additional technique based on the accessibility of DNA to a methyltransferase.

MNase-seq. MNase is a secreted glycoprotein with a preference for single-stranded DNA and RNA. Mechanistically, it is thought to cleave one strand of DNA when the helix 'breathes' and subsequently cleave the other strand to generate a double-strand break. It then 'nibbles' the exposed DNA end until it reaches an obstruction, such as a nucleosome[24,25]. The positions of not only nucleosomes but also paused RNA polymerase II (Pol II) could be mapped using MNase digestion followed by sequencing (MNase-seq) (FIG. 1a) in *Drosophila melanogaster* cells[26], and this has been confirmed by Pol II ChIP[27]. MNase-seq with paired-end sequencing has also been used to map the positions of both nucleosomes and sequence-specific TFs in yeast[28]. These studies showed that MNase can map fragments as short as ~50 bp, which correspond to non-nucleosomal particles. To maximize the amount of information that could be obtained, MNase mapping has been combined with a modified sequencing library construction protocol that enables the recovery and sequencing of fragments ranging in size from ~25 bp to 500 bp[29]. Using this approach, the chromatin structures around binding sites were delineated for >100 yeast TFs at base-pair resolution as fragment midpoint-versus-length plots (V-plots)[29]. MNase-seq has also been used to characterize yeast centromeres[30] and the *D. melanogaster* epigenome in response to heat shock[27].

MNase-seq maps both nucleosomes and non-histone proteins in a single experiment, making it a cost-effective method for epigenomic profiling (TABLE 1). Additionally, it does not rely on affinity reagents and is therefore applicable to epigenomic features for which such reagents are not available. MNase-seq with paired-end sequencing also gives single-base resolution of protein binding sites when used in combination with V-plotting[29]. A basic limitation of MNase-seq is that the identity of non-nucleosomal particles cannot be conclusively determined. However, as discussed in the section on ChIP below, MNase-digested, uncrosslinked chromatin can be used for ChIP–seq. MNase cleavage is also biased towards AT-rich regions, which might influence MNase-based epigenomic maps[31,32], although this bias seems to be small in practice[33] and can be computationally alleviated if necessary[34].

DNase-seq. DNase I is a nonspecific endonuclease that acts on single- and double-stranded DNA. Similarly to FAIRE-seq and Sono-seq, mapping of the ends of DNase I-released fragments to a reference genome identifies regions of chromatin accessibility[19]. The combination of DNase I mapping and sequencing as digital genomic footprinting (DGF)[35] (FIG. 1a; TABLE 1) enables single-base resolution mapping of TF binding sites.

More recently, a modification of DNase-seq termed DNase I-released fragment length analysis of hypersensitivity (DNase-FLASH) was introduced[36]. This approach involves sequencing of different size ranges of DNase I-digested chromatin to footprint both nucleosomes and TFs. Similarly to MNase-seq, DNase-FLASH requires paired-end sequencing to determine precise fragment lengths, and the resulting data can be used to generate V-plots. As with MNase-seq, DNase-FLASH provides information about nucleosomes and non-histone proteins from a single experiment and does not rely on affinity reagents. However, also like MNase, DNase I suffers from cleavage bias, which can lead to identification of artefactual TF footprints[37], but this effect can be corrected computationally. An additional drawback of DNase-seq is that very high sequencing depth is necessary to comprehensively map TF footprints in large genomes[38,39]: many DNase-hypersensitive regions do not reach saturation even with >500 million uniquely mapping DNase I cleavages[39]. Furthermore, DNase I is inhibited by high concentrations of actin[40], and DGF therefore requires chromatin purification or an alternative endonuclease for epigenomic profiling in tissues[41].

ATAC-seq. Tn5 transposase catalyses the transfer of transposons from one genomic location to another through a 'cut-and-paste' mechanism. Treatment of genomic DNA with a hyperactive form of Tn5 transposase is often used to create sequencing libraries. DNA fragmentation and ligation of sequencing adapters are carried out in a single step known as 'tagmentation', which enables simple, rapid library preparation[42]. A recent study exploited the preferential integration of transposons into nucleosome-depleted regions to interrogate chromatin structure *in vivo*[43]. This method, termed assay

---

Digital epigenomic analysis
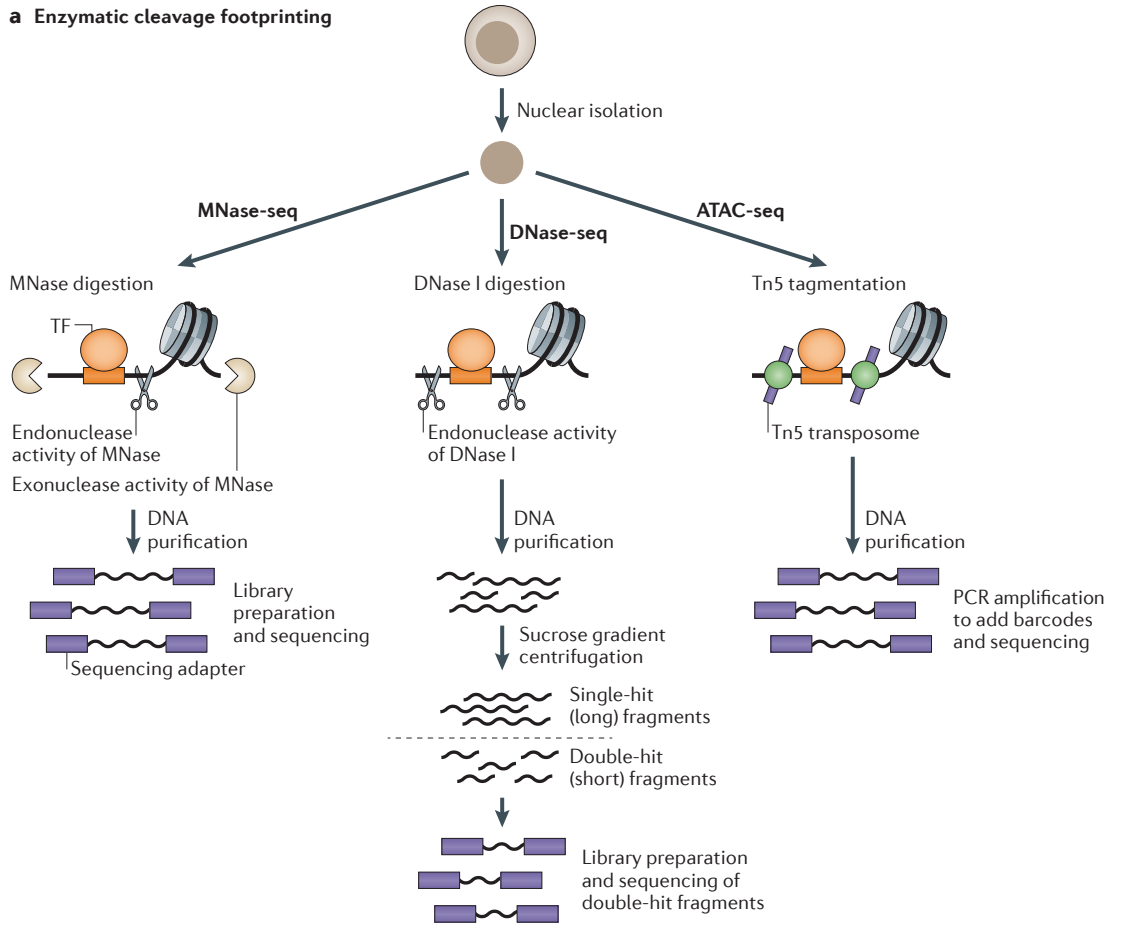The use of methods with sequencing-based readouts to interrogate the epigenome.

Sequencing library
A collection of DNA fragments prepared for high-throughput sequencing by the addition of specific adapter sequences.

Fragment midpoint-versus-length plots
(V-plots). Representations of paired-end sequencing data in which a point corresponding to the midpoint of a paired-end read is plotted in two-dimensional space. The *x* coordinate of the point represents the distance of the read midpoint from a defined genomic feature, and its *y* coordinate represents the length of the fragment from which it was derived.
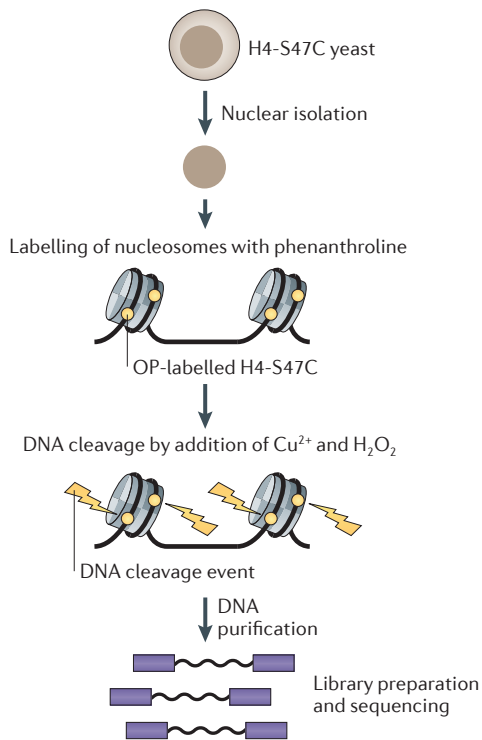
Affinity reagents
Antibodies or other molecules used to recover specific proteins from a complex mixture.

Tagmentation
Simultaneous fragmentation and incorporation of sequencing adapters into chromatin using Tn5 transposase.
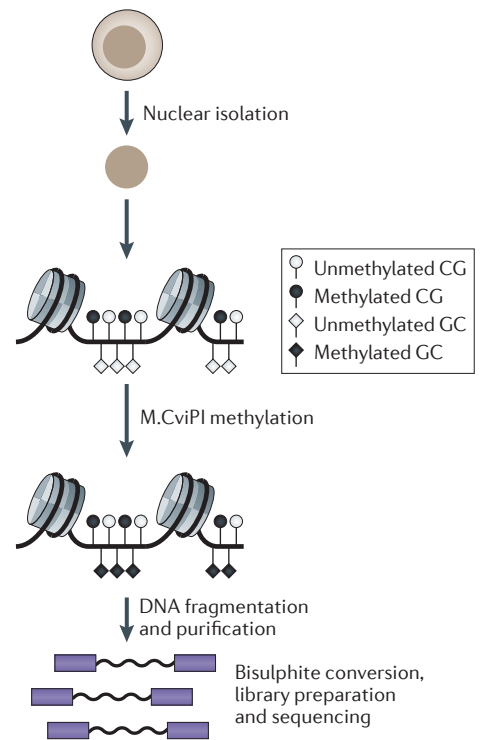
a Enzymatic cleavage footprinting

MNase-seq

DNase-seq

ATAC-seq

Nuclear isolation

MNase digestion

DNase I digestion

Tn5 tagmentation

TF

Endonuclease activity of MNase

Exonuclease activity of MNase

Endonuclease activity of DNase I

Tn5 transposome

DNA purification

DNA purification

DNA purification

Library preparation and sequencing

Sequencing adapter

Sucrose gradient centrifugation

PCR amplification to add barcodes and sequencing

Single-hit (long) fragments

Double-hit (short) fragments

Library preparation and sequencing of double-hit fragments

b Chemical cleavage footprinting

H4-S47C yeast

Nuclear isolation

Labelling of nucleosomes with phenanthroline

OP-labelled H4-S47C

DNA cleavage by addition of Cu^{2+} and H_2O_2

DNA cleavage event

DNA purification

Library preparation and sequencing

c DNA methylation footprinting

Nuclear isolation

○ Unmethylated CG
● Methylated CG
◇ Unmethylated GC
◆ Methylated GC

M.CviPI methylation

DNA fragmentation and purification

Bisulphite conversion, library preparation and sequencing

◀ Figure 1 | **Methods for chromatin footprinting. a** | An overview of experimental footprinting methods based on enzymatic cleavage is shown. For micrococcal nuclease digestion followed by sequencing (MNase-seq), nuclei are isolated and treated with MNase to digest chromatin, and DNA is purified for sequencing (left). For deoxyribonuclease I digestion followed by sequencing (DNase-seq), nuclei are isolated and chromatin is digested with DNase I (middle). DNA is purified and subjected to sucrose gradient centrifugation to isolate short 'double-hit' fragments (that is, fragments released by two closely spaced DNase I cleavages), which are then sequenced. For assay for transposase-accessible chromatin using sequencing (ATAC-seq), isolated nuclei are treated with Tn5 transposomes loaded with sequencing adapters, which both fragment and tag chromatin in a process known as 'tagmentation' (right). Tagmented DNA is then purified, amplified by PCR and sequenced. **b** | In targeted chemical cleavage, nuclei from yeast harbouring a single copy of histone H4 with the S47C substitution (H4-S47C) are labelled with (N-(1,10-phenanthroline-5-yl)iodoacetamide) (OP) and treated with Cu^{2+} ions and hydrogen peroxide (H_2O_2) to cleave DNA. DNA is then isolated and sequenced. **c** | In nucleosome occupancy and methylation sequencing (NOMe-seq), isolated nuclei are treated with M.CviPI, which methylates cytosines in the GC dinucleotide context. DNA is fragmented, purified and subjected to bisulphite conversion. Bisulphite-treated DNA is then sequenced. TF, transcription factor.

for transposase-accessible chromatin using sequencing (ATAC-seq) (FIG. 1a; TABLE 1), generated results comparable to those of other chromatin profiling methods such as FAIRE-seq and DNase-seq in terms of chromatin accessibility[44]. Notably, these results were obtained with only 50,000 cells, which is several orders of magnitude less than the standard input requirements for FAIRE-seq and DNase-seq. High-quality ATAC-seq data were also obtained with as few as 500 cells, although at a cost to sensitivity. Similarly to DNase-FLASH and MNase-seq, ATAC-seq also provides information regarding nucleosome occupancy around regulatory regions and footprinted TFs.

The low input requirement of ATAC-seq suggests that it is uniquely applicable to precious samples such as rare developmental cell types and patient samples. Furthermore, the ease and rapidity of the ATAC-seq protocol may be compatible with clinical epigenomics. Indeed, a protocol has been developed that takes 275 minutes from blood draw to the time of sequencing. A factor that may limit the resolution of ATAC-seq is steric hindrance between adjacent Tn5 molecules, which gives rise to a minimum spacing of ~38 bp between transposition events[42]. This would provide an explanation for the propensity of ATAC-seq to give information primarily about nucleosomes proximal to regulatory regions, which tend to be nucleosome-depleted and thus provide ample DNA for Tn5 association. This is in contrast to MNase-seq, which assays all nucleosomes. Tn5 thus avoids the nucleosome-dense inactive genome and greatly reduces the sequencing depth required for a genome-wide map.

Targeted chemical cleavage. To precisely map nucleosome centres across the genome, a targeted chemical cleavage approach was developed[16] (FIG. 1b; TABLE 1). For this approach, a yeast strain was created that carried a single copy of histone H4 bearing a serine-to-cysteine substitution at amino acid 47 (H4-S47C), which is in close proximity to DNA at the nucleosome centre (dyad). This mutation enables covalent attachment

of the suphydryl-reactive, copper-chelating label N-(1,10-phenanthroline-5-yl)iodoacetamide. Yeast cells are permeabilized, cysteine residues are labelled, and cells are treated with Cu^{2+} ions and hydrogen peroxide, which catalyse the formation of hydroxyl radicals that cleave DNA at specific positions relative to the nucleosome dyad[45]. This approach yields precise maps of nucleosome occupancy and positioning, and can be used to characterize non-canonical nucleosome structures[45]. Chemical cleavage has also been used to analyse DNA binding of the bacterial Trp repressor and Fis nucleoid-associated protein, as well as *D. melanogaster* homeobox TFs[46–48], indicating that it is applicable to the analysis of non-histone proteins; however, this requires cysteine mutation of DNA-contacting residues in the DNA-binding domain of a protein, which could conceivably lead to structural alteration of the DNA-binding domain and compromise the protein's ability to bind to DNA.

DNA methylation footprinting. Cleavage-based methods of chromatin footprinting damage DNA and potentially disrupt chromatin structure. A gentle alternative is DNA methylation footprinting (FIG. 1c). This method does not damage DNA and leaves relatively innocuous methyl groups, causing minimal disruption to chromatin. Indeed, such methods can even be carried out *in vivo*. For example, DNA adenine methyltransferase identification (DamID), which involves fusion of a GATC methyltransferase to a factor of interest, has been used to determine genomic distributions of both TFs and histones *in vivo*[49,50]. Chromatin footprinting of CG and GC dinucleotides *in vivo* using the SssI and CviPI methyltransferases (M.SssI and M.CviPI) was also combined with bisulphite sequencing as a readout for DNA methylation, the rationale being that methylation levels would be correlated with the accessibility of the surrounding DNA[51]. More recently, methylation footprinting with M.CviPI has been carried out on a genome-wide scale as nucleosome occupancy and methylation sequencing (NOMe-seq)[52] (TABLE 1). NOMe-seq assays nucleosome occupancy and positioning through methylation of GC dinucleotides by M.CviPI. The use of M.CviPI allows distinction between ectopic methylation used for nucleosome footprinting and endogenous methylation at CG dinucleotides following bisulphite conversion of DNA and sequencing. Similar to other footprinting methods, NOME-seq does not conclusively identify the particle responsible for protection and additionally relies on the presence of GC nucleotides. It provides rich data by simultaneously interrogating two distinct features of the epigenome and is thus particularly useful for organisms with CG DNA methylation.

Chromatin immunoprecipitation

ChIP is used to determine binding sites for specific proteins and, in conjunction with formaldehyde crosslinking (X-ChIP), is one of the most widely used epigenomic techniques. In combination with tiled microarray analysis (ChIP–chip) or massively parallel short-read sequencing (ChIP–seq), ChIP has been

REVIEWS

Table 1 | **An overview of digital epigenomic techniques**

Method	Features	Refs
MNase-seq	• Base-pair resolution • Does not rely on affinity reagents • Maps nucleosomes and non-histone proteins simultaneously	29
DNase-seq	• Base-pair resolution • Does not rely on affinity reagents • Maps nucleosomes and non-histone proteins simultaneously in DNase-FLASH • Requires high sequencing depth to footprint large genomes • Efficiency is reduced by actin contamination	35,36
ATAC-seq	• Base-pair resolution at sites of accessible chromatin • Does not rely on affinity reagents • Maps nucleosomes and non-histone proteins simultaneously • Simple and rapid protocol • Low input requirement	44
Targeted chemical cleavage	• Base-pair resolution of nucleosome centres • Does not rely on affinity reagents • Can map nucleosome substructures	16
NOMe-seq	Simultaneously maps DNA methylation and nucleosomes	52
ChIP-exo	• Base-pair resolution • High input requirement	57
High-resolution X-ChIP–seq	• Base-pair resolution • Complete solubilization of certain chromatin proteins	63
ORGANIC	• Base-pair resolution • No crosslinking artefacts • Potentially poor solubilization of proteins	64,65
Histone variant (H3.3) profiling	• Identifies sites of histone turnover • Provides limited information on turnover dynamics	70,71
Tagged protein replacement	• Some temporal resolution of chromatin binding dynamics • Requires tagged alleles	75
CATCH-IT	Does not require epitope-tagged histones	81
Crosslinking kinetic analysis	Very high temporal resolution	88
NET-seq	• Base-pair resolution • Maps Pol II in unperturbed cells • Requires solubilization of the Pol II complex	93
PRO-seq	• Base-pair resolution • Requires extensive nuclear manipulation • Maps elongation-competent RNA polymerases	96
3′NT method	• Base-pair resolution • Maps total RNA polymerases	98

3′NT, 3′ nascent transcript; ATAC-seq, assay for transposase-accessible chromatin using sequencing; CATCH-IT, covalent attachment of tags to capture histones and identify turnover; ChIP-exo, chromatin immunoprecipitation and exonuclease digestion; DNase-FLASH, deoxyribonuclease I-released fragment length analysis of hypersensitivity; DNase-seq, DNase I digestion followed by sequencing; MNase-seq, micrococcal nuclease digestion followed by sequencing; NET-seq, native elongating transcript sequencing; NOMe-seq, nucleosome occupancy and methylation sequencing; ORGANIC, occupied regions of genomes from affinity-purified naturally isolated chromatin; Pol II, RNA polymerase II; PRO-seq, precision nuclear run-on and sequencing; X-ChIP–seq, crosslinking ChIP followed by sequencing.

used to study the genomic distributions of hundreds of proteins[53–55]. Although many important biological insights have been gained through the use of ChIP–chip and ChIP–seq, these methods do have limitations (BOX 1). In particular, the resolution of standard ChIP–seq is limited by the use of sonication to fragment and solubilize chromatin. Sonication produces a heterogeneous mixture of fragment sizes that preclude

precise determination of protein binding sites, and this problem is exacerbated by the selection of fragments that are 200–400 bp in length for sequencing, which is a common practice in ChIP–seq library preparation protocols[56]. Most ChIP–seq libraries are also sequenced from one end only, and the resulting reads are computationally extended to an approximate fragment size. Together, these issues have limited the resolution of genome-wide ChIP. However, several recent techniques have combined enzymatic digestion of chromatin with ChIP to improve its resolution.

ChIP-exo. ChIP and exonuclease digestion (ChIP-exo) is a high-resolution modification of ChIP–seq[57] (FIG. 2a; TABLE 1) in which X-ChIP is performed as usual, but immunoprecipitated chromatin is digested with λ-exonuclease. λ-exonuclease processively degrades naked DNA in the 5′-to-3′ direction until it reaches a protein–DNA crosslink. Resected DNA is then prepared for sequencing through a specialized library construction protocol, and the 5′ ends of the resulting reads are mapped back to a reference genome, resulting in pile-ups of reads (peaks) that correspond to the locations of protein–DNA crosslinks at single-base resolution. Pairs of closely spaced peaks, which represent barriers to λ-exonuclease digestion, are taken to represent protein binding sites. ChIP-exo has been applied to several yeast TFs[57], the human insulator-binding protein CCCTC-binding factor (CTCF)[57,58], the yeast chromatin remodeller Isw2 (REF. 59) and the mouse TF sex-determining region Y-box 2 (SOX2)[60], demonstrating its use for mapping a variety of proteins. The resolution of ChIP-exo is a vast improvement over standard sonication-based ChIP and provides very high signal-to-noise ratios (300–2,800-fold higher than that of standard ChIP–seq). Analysis of ChIP-exo data is less straightforward than that of standard ChIP–seq data[61], although multiple new software packages, including GEM[61], MACE and CexoR, have dedicated functionality for ChIP-exo data analysis[62]. ChIP-exo is also noted for having a particularly high input requirement[9]. As with any formaldehyde crosslinking-based method, ChIP-exo is potentially susceptible to crosslinking artefacts (BOX 2).

High-resolution X-ChIP–seq. In a modification of standard X-ChIP–seq, crosslinked cells are lysed and chromatin is digested with MNase, sonicated and immunoprecipitated[63] (FIG. 2b; TABLE 1). Similarly to ChIP-exo, protein–DNA crosslinks will act as barriers to nuclease processivity, leaving protein-bound DNA intact. Immunoprecipitated DNA is then purified and subjected to paired-end sequencing using a modified library preparation protocol that allows sequencing of a broad range of fragment sizes and high-resolution determination of the binding sites of a protein. The use of both MNase and sonication results in near-complete solubilization of some chromatin-bound proteins, making it especially useful for large complexes that resist solubilization, such as the Pol II holoenzyme[63]. However, similarly to standard X-ChIP and ChIP-exo, high-resolution X-ChIP–seq may suffer from crosslinking artefacts (BOX 2).

Native ChIP
Chromatin immunoprecipitation (ChIP) using chromatin that has not been crosslinked with formaldehyde or any other crosslinking agent.

ORGANIC. Native ChIP can both improve the resolution of ChIP and circumvent issues associated with formaldehyde crosslinking (BOX 2). In the occupied regions of genomes from affinity-purified naturally isolated chromatin (ORGANIC) method, soluble chromatin extracted from MNase treatment of nuclei[29] is used as an input for ChIP–seq[64,65] (FIG. 2c; TABLE 1). Thus, the input sample for ORGANIC provides a genome-wide footprinting of factors, and ChIP pulldown provides a factor-specific map in a single experiment using a simple library preparation protocol. This method was first applied to ATP-dependent DNA translocases to address questions regarding their *in vivo* actions[64,66]. ORGANIC profiling of yeast and *D. melanogaster* TFs revealed that this method is highly sensitive and specific, identifying more binding sites with consensus motifs than previous X-ChIP studies[65]. Most recently, ORGANIC has been used to profile *D. melanogaster* Polycomb group proteins[67]. ORGANIC may offer solutions to issues associated with formaldehyde crosslinking, although there are other issues, such as protein solubility, that should be considered when designing an ORGANIC experiment (BOX 2).

Box 1 | Validation of transcription factor binding sites

The presence of a characteristic sequence motif within a ChIP–seq (chromatin immunoprecipitation followed by high-throughput sequencing) peak is traditionally considered to be the 'gold standard' for validation of a factor-bound region[65,119,120]. In bacteria, motifs determined by *in vitro* experiments predict *in vivo* binding with a near-perfect correspondence between transcription factor (TF) occupancy and canonical motif observed in many cases[121]. Binding of TFs to non-canonical sites is also observed, but these binding events are generally viewed with some scepticism and require additional validation to determine their biological relevance[121].

   A major result of numerous genome-wide mapping studies in complex eukaryotes has been the lack of consensus motifs within ChIP–seq-detected peaks for many sequence-specific TFs. This phenomenon is exemplified by recent results from the Encyclopedia of DNA Elements (ENCODE) Consortium, which indicated that, for a group of 36 TFs with known DNA-binding specificities, ~36–100% of their binding sites did not contain the expected motif (or motifs)[39]. This result brought into sharp relief an important question posed by many eukaryotic ChIP–seq studies: given the exquisite sequence specificities of TFs evolved over millions of years, why would such factors bind predominantly to regions lacking their cognate binding sites? One explanation is that these sites represent indirect binding of TFs to chromatin; that is, they are tethered to DNA by one or more additional factors that are directly associated with the DNA. Indirect recruitment of TFs to chromatin is a well-established phenomenon[122] and could account for the lack of consensus motifs at these sites. However, without experimental validation of these putative indirect interactions, there remains the possibility that these are crosslinking artefacts. Many TFs are expressed at levels that are expected to thermodynamically drive nonspecific interactions with chromatin[123,124], and the majority of TF–genome interactions in eukaryotes are thought to be such opportunistic interactions[124]. For example, the TF sex-determining region Y-box 2 (SOX2) scans the genome for binding sites and samples many genomic regions prior to locating a consensus motif. ChIP and exonuclease digestion (ChIP-exo) of SOX2 revealed numerous low-occupancy peaks with degenerate motifs, which indicates that capture of TFs sampling low-affinity motifs might contribute to artefactual ChIP signals[60]. Indeed, sites of low TF occupancy in the *Drosophila melanogaster* embryo do not drive reporter expression[124]. Combined with the prevalence of formaldehyde-reactive primary amines in histones and other chromatin-associated proteins, such artefactual interactions might be captured through protein–protein crosslinks. Highlighting the need for rigorous validation of ChIP–seq experiments, a recent study found that ~45% of ChIP–seq data sets deposited in the Gene Expression Omnibus (GEO) as of 1 April 2012 were of poor or intermediate quality[125].

Chromatin dynamics

The epigenome is a highly dynamic entity[68]. Nucleosomes are slid, evicted and replaced, and TFs and other non-histone proteins dynamically interact with the genome on a timescale of seconds[60]. Although footprinting and ChIP techniques provide valuable information about locations of protein binding, they yield limited data regarding dynamics. We discuss below various techniques designed to give insights into the dynamic nature of the epigenome.

Histone variant profiling. Initial efforts to characterize chromatin dynamics focused on the histone variant H3.3, which is incorporated into chromatin independently of DNA replication[69] and thus marks regions of nucleosome turnover (FIG. 3a; TABLE 1). ChIP–chip profiling of biotinylated H3.3 in *D. melanogaster* S2 cells revealed enrichment over active genes and regulatory regions, such as binding sites for Trithorax group and Polycomb group proteins[70,71], which is consistent with cycles of nucleosome disruption and binding of Pol II and TFs at these locations. Similar results have been obtained in mammalian cells[72,73] and *Caenorhabditis elegans* embryos[74]. Although genomic profiling of H3.3 enrichment provides information about the location at which nucleosome turnover occurs, it gives limited information regarding turnover kinetics and requires epitope tagging of H3.3.

Tagged protein replacement. The next technique introduced to study epigenome dynamics was tagged protein replacement. In this method, one allele of an epitope-tagged protein of interest is constitutively expressed, and a second allele tagged with a different epitope is controlled by an inducible promoter. Upon induction, a time series of genome-wide ChIP experiments is carried out with antibodies directed against each epitope tag (FIG. 3b; TABLE 1). For each time point, the ratio of induced to constitutively expressed protein at each location of interest is used to calculate a rate of turnover. This approach was first used to assess histone turnover in budding yeast[75,76] and has also been used to assay chromatin interaction dynamics of the basal initiation factor TATA-box binding protein (TBP)[77] and the general regulatory TF Rap1 (REF. 78) in yeast, H3.3 dynamics in mouse cells[79], and histone turnover and splitting in human cells[80]. Tagged protein replacement is limited by the need for epitope-tagged alleles, which are easy to generate in yeast but more difficult in multicellular eukaryotes, which require episomal vectors or viral targeting, although rapid advances in genome editing technologies such as CRISPR–Cas9 (clustered regularly interspaced short palindromic repeat–CRISPR-associated protein 9) may facilitate epitope tagging of endogenous alleles in less genetically tractable organisms. Additionally, although tagged protein replacement offers greater resolution of genome-wide chromatin interaction dynamics than ChIP alone, induction of the replacement protein requires 20–45 minutes in yeast[75,77,78] and up to several hours in mouse cells[79], which limits the temporal resolution of the technique.

CATCH-IT. A metabolic labelling approach, termed covalent attachment of tags to capture histones and identify turnover (CATCH-IT), yields high temporal resolution and circumvents the need for genetic transformation to map nucleosome dynamics[81] (FIG. 3c; TABLE 1). In CATCH-IT, cells are starved of methionine and treated with the methionine analogue L-azido-homoalanine (AHA). Following incubation to allow co-translational incorporation of AHA into proteins, AHA-containing proteins are biotinylated through a cycloaddition reaction between the azide group of AHA and a biotin-linked alkyne. Chromatin is then

digested with MNase, and biotinylated proteins are purified with streptavidin. Non-histone proteins and H2A–H2B dimers are removed through a series of stringent washes, and $(H3-H4)_2$ tetramer-associated DNA is then analysed by microarray hybridization or high-throughput sequencing. The amount of DNA recovered at a particular locus by streptavidin pulldown is thus proportional to the level of histone turnover at that region. Indeed, in *D. melanogaster* S2 cells, CATCH-IT signal is well correlated with histone H3.3 enrichment, which is consistent with increased incorporation of newly synthesized, AHA-containing histones at sites
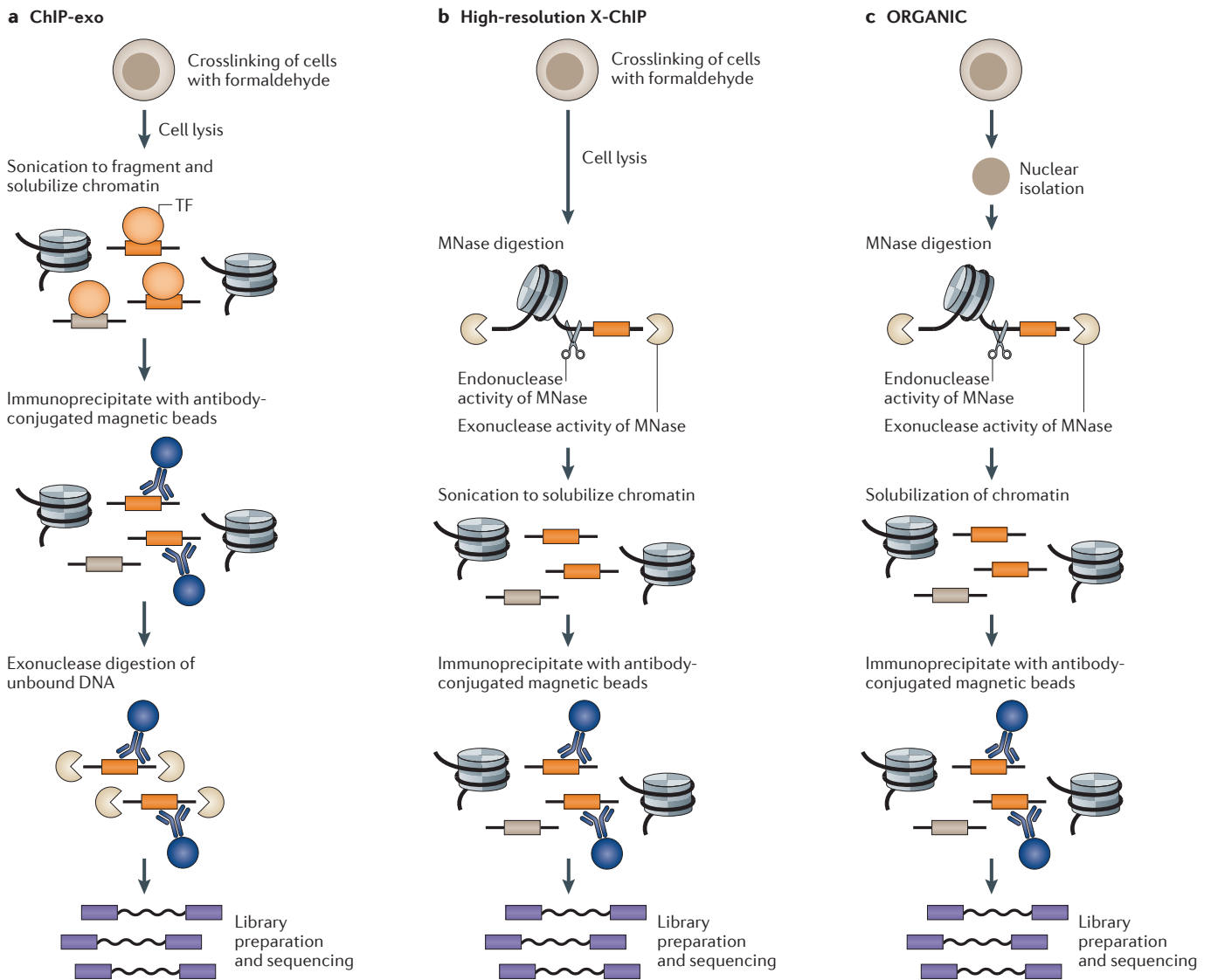

Figure 2 | **High-resolution ChIP methodologies.** **a** | In chromatin immunoprecipitation and exonuclease digestion (ChIP-exo), crosslinked cells are sonicated to fragment and solubilize chromatin. ChIP is then carried out with an antibody directed against the protein of interest. Immunoprecipitated DNA is digested with exonucleases to remove DNA that is not protected by the protein; resected DNA is then purified and sequenced. **b** | In high-resolution crosslinking ChIP (X-ChIP), crosslinked cells are lysed and chromatin is digested with micrococcal nuclease (MNase). Chromatin is then sonicated to improve solubility. An antibody directed against the protein of interest is then used to immunoprecipitate DNA, which is then purified and sequenced. **c** | Similar to MNase-seq, in occupied regions of genomes from affinity-purified naturally isolated chromatin (ORGANIC), nuclei are isolated and chromatin is digested with MNase and solubilized. An antibody directed against the protein of interest is then used to immunoprecipitate DNA, which is then purified and sequenced. TF, transcription factor.

of high histone turnover[81]. CATCH-IT was also used to probe the effects of heat shock[27] and anthracycline anticancer drugs[82] on nucleosome turnover.

Torsion profiling. *In vitro* evidence suggests that negative and positive DNA supercoiling have opposite effects on nucleosome formation[83] and might thus affect nucleosome dynamics *in vivo*. Most *in vivo* methods for measuring supercoiling use psoralen derivatives such as trimethylpsoralen (TMP), which intercalate into DNA and can be crosslinked to opposite strands using ultraviolet irradiation[84]. More TMP intercalates into underwound (that is, negatively twisted) DNA than into DNA that is not under torsional stress and vice versa for overwound (that is, positively twisted) DNA, so that the amount of TMP-crosslinked DNA in a region is a measure of DNA torsion. Various methods have been used to purify TMP-crosslinked DNA, such as immunoprecipitation of tagged TMP–DNA complexes[85], thermal denaturation of DNA followed by gel electrophoretic isolation of the 'snap-back' fraction (that is, the fraction of DNA that, after denaturation, returns to a double-stranded

state owing to the presence of psoralen crosslinks)[86] and exonucleolytic digestion to resect DNA back to sites of TMP–DNA crosslinking[84,87].

Crosslinking kinetic analysis. Analysis of the dynamic interactions between TFs and chromatin *in vivo* is a difficult problem. ChIP approaches generally provide accurate information as to where proteins bind to chromatin but little information on interaction dynamics. For example, a robust ChIP signal could indicate either a stable long-lived interaction or repeated capture of transient interactions due to the long formaldehyde crosslinking step used in most standard X-ChIP protocols, and it is thus impossible to distinguish legitimate signals from artefacts or to infer relative occupancy and dynamics. Conversely, live-cell imaging approaches provide precise information about dynamics but not high-resolution data regarding binding locations. Tagged protein replacement has been used to measure chromatin interaction dynamics, but the time required to generate the competitor species severely limits its temporal resolution[75,77–79]. To address chromatin binding dynamics on the seconds-to-minutes timescale and to avoid issues with binding site saturation due to extended crosslinking, the crosslinking kinetic analysis[88] approach was introduced (FIG. 3e; TABLE 1). This method uses a mathematical model based on kinetic principles that describes the relationship between ChIP signal and formaldehyde crosslinking times. From this, on rate, off rate and the fraction of bound sites at steady state can be extracted. Several assumptions of this model were confirmed using the yeast TF Gal4. Crosslinking kinetic data for two additional TFs were consistent with fluorescence recovery after photobleaching (FRAP) data, a live-cell imaging approach for the analysis of protein dynamics. Although crosslinking kinetic analysis offers high temporal resolution, specialized quench-flow equipment is necessary for accurate crosslinking and quenching at the shortest timescales; however, at time points greater than ~6 seconds these can be performed by hand. As with any formaldehyde-based method, crosslinking kinetic analysis may be susceptible to crosslinking artefacts (BOX 2), although the brief timescales used can probably minimize this concern.

The epigenome and transcription

Transcription occurs on a chromatin template, and it is therefore important to understand how the epigenome influences transcription. The relationship between transcription and the nucleosome has been extensively studied *in vitro*[89], and nucleosomes are known to be physical barriers to transcription that may cause stalling, backtracking and arrest of Pol II. However, the situation *in vivo* is much more complex, with many protein complexes influencing chromatin structure. *In vivo*, the relationship between Pol II and chromatin has been studied mostly using Pol II ChIP–seq[90]. However, the previously discussed limitations of ChIP–seq, particularly its low spatial resolution, have hampered efforts to map Pol II genome-wide, and these limitations have been confounded by the insolubility of engaged Pol II, which makes full recovery difficult. However, several

---

Box 2 | **Is crosslinking necessary?**

Epigenomic methods are performed either with crosslinking or natively. We focus below on the limitations of crosslinking chromatin immunoprecipitation (X-ChIP) and native ChIP, although the points discussed are applicable to other epigenomic methodologies.

Although X-ChIP is perhaps the most widely used epigenomic method, caution may be warranted when interpreting X-ChIP results. In yeast, proteins that are not associated with transcriptional activity, such as the centromeric histone variant Cse4, could be detected at the promoters and bodies of highly expressed genes[126]. Similar results were also observed with proteins that are not expected to associate with the genome, including nuclear localized GFP and the Golgi enzyme Mnn10 (REF. 127). Possible artefacts have also plagued epigenomic analysis of Mediator[128,129]. Although these artefacts might be due to technical issues, they may also be due to crosslinking, as high concentrations of transcription factors (TFs) in nuclei are expected to drive nonspecific interactions with chromatin through mass action[123,124]. As formaldehyde preferentially generates protein–protein crosslinks[116,117], many of these opportunistic interactions may be captured through protein–protein crosslinking, particularly with nucleosomes, which are rich in formaldehyde-reactive primary amines. Ultraviolet crosslinking, which generates only protein–nucleic acid crosslinks, may offer a solution to artefactual signals generated by protein–protein crosslinking[130]. X-ChIP experiments might also show a bias towards regions of accessible chromatin[23,65,125,131,132], which are frequently sampled during TF binding site scanning[60]. Although recent studies in yeast have attempted to compensate for this by performing normalization to input or no-tag controls, these were generally found to be inadequate[126,127]. Normalization is not carried out for large genomes, as input samples must be sequenced at sufficient depth to provide whole-genome coverage.

Native ChIP is performed without crosslinking. It is usually applied to nucleosomes, the assumption being that the wrapping of DNA around histones precludes rearrangement during chromatin preparation and immunoprecipitation. It is also often assumed that native ChIP is unsuitable for profiling non-histone proteins owing to potential rearrangement. This was addressed by ORGANIC (occupied regions of genomes from affinity-purified naturally isolated chromatin) profiling of the yeast general regulatory TF Reb1 from a mixed sample of yeast and *Drosophila melanogaster* nuclei. Immunoprecipitated Reb1 is robustly enriched in yeast but not *D. melanogaster* DNA, which argues against rearrangement. Moreover, sites for Reb1 and a second yeast general regulatory TF, Abf1, showed robust DNase footprints, indicating occupancy[65]. Solubility of proteins of interest can also be an issue with native ChIP: as harsh detergents and sonication are not used, recovery might be lower than for X-ChIP, especially for large complexes, in which case high-resolution X-ChIP[63] is preferred.

---

a Histone variant profiling

High turnover   Low turnover

b Tagged protein replacement

High turnover   Low turnover

c CATCH-IT

Cells starved of methionine

Cells + AHA

Isolation of nuclei

Cycloaddition reaction

Exonuclease activity of MNase — Biotin
Endonuclease activity of MNase

Removal of H2A–H2B dimers and non-histone proteins

Affinity purification of biotinylated histones, DNA purification and sequencing

d Torsion profiling

Negative   Positive
Torsional stress   Torsional stress

TMP

UV crosslinking

Exonuclease digestion

Library preparation and sequencing

e Crosslinking kinetic analysis

High on-rate   Low on-rate
TF

Time

Crosslink   TF binding site

Crosslinking ChIP time course

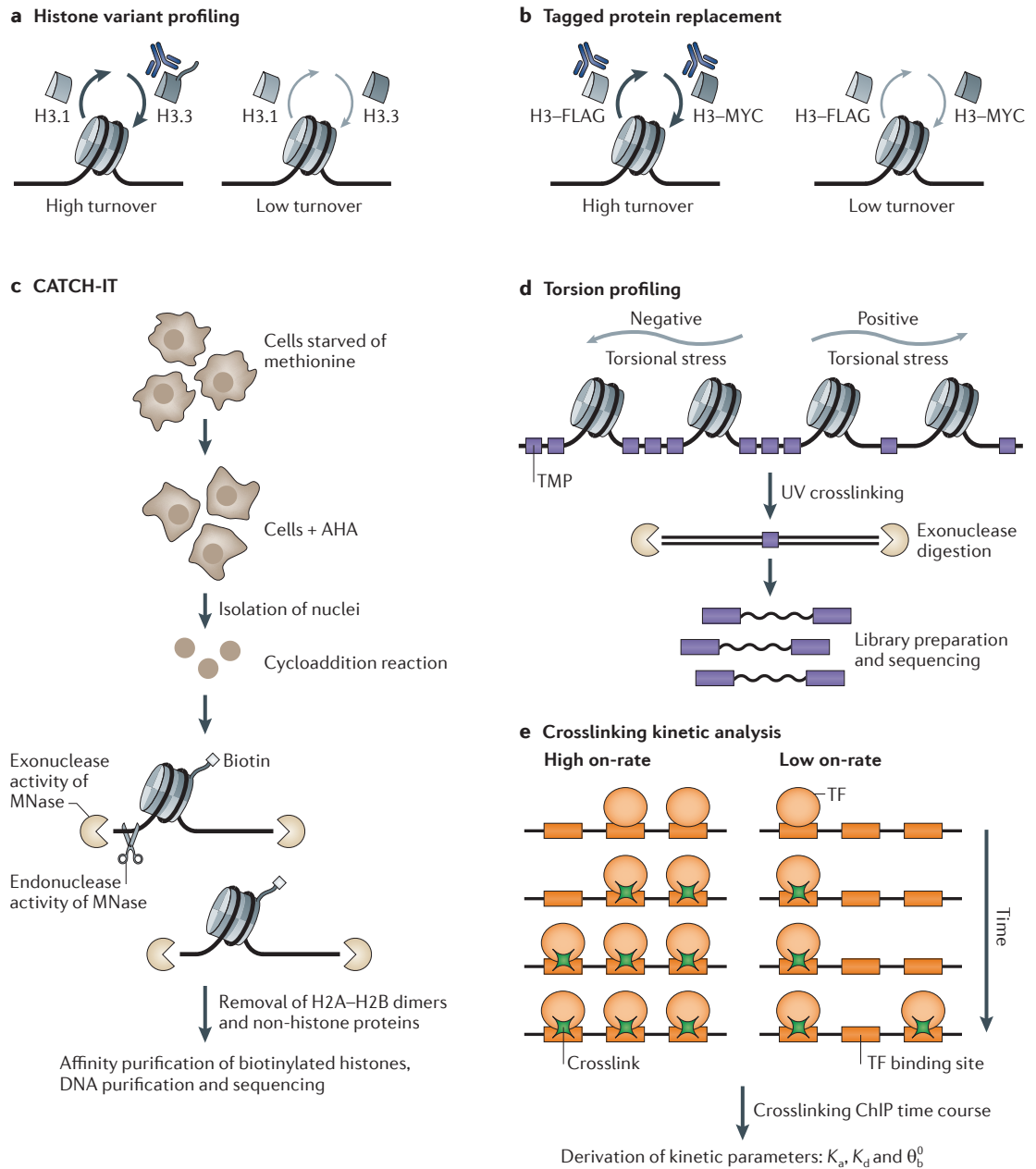Derivation of kinetic parameters: K_a, K_d and θ_b^0

Figure 3 | **Methods for analysing chromatin dynamics. a** | In histone variant profiling, histone H3.3 incorporation is highest at regions of rapid replication-independent nucleosome turnover, and affinity purification of H3.3 thus gives spatial and limited temporal information about this phenomenon. **b** | In tagged protein replacement, a histone or transcription factor (TF) of interest bearing an epitope tag (in this case, FLAG) is constitutively expressed. A second copy of the protein bearing a different epitope (in this case, MYC) is then induced. Time-course affinity purification of the protein with each epitope tag then gives spatial and limited temporal information about the turnover of histones or other proteins. **c** | In covalent attachment of tags to capture histones and identify turnover (CATCH-IT), cells are starved of methionine and labelled with the methionine analogue ʟ-azidohomoalanine (AHA). AHA-labelled proteins are biotinylated by cycloaddition, chromatin is fragmented with micrococcal nuclease (MNase), and biotinylated proteins are purified with streptavidin. H2A–H2B dimers and non-histone proteins are washed away, and DNA is purified and sequenced. **d** | In torsion profiling, trimethylpsoralen (TMP) inserts into DNA and can be crosslinked to opposite strands using ultraviolet (UV) irradiation. The amount of TMP-crosslinked DNA in a region is a measure of DNA torsion. **e** | In crosslinking kinetic analysis, different classes of TF binding sites are examined. The sites on the left have a high on-rate for a given TF, whereas the sites on the right have a low on-rate for a particular TF. The time course illustrates how occupancy of each site within the cell population is predicted to change with formaldehyde crosslinking. Chromatin immunoprecipitation (ChIP) with an antibody directed against a factor of interest is carried out after various lengths of formaldehyde crosslinking, and the ChIP signal is used to derive kinetic parameters — such as the rate of association (K_a), the rate of dissociation (K_d) and the fraction of bound sites at steady state (θ_b^0) — at each site.

alternative techniques have recently been developed to address *in vivo* the relationship of transcription to the chromatin template at high resolution (see below).

Nascent RNA-seq. High-throughput RNA sequencing (RNA-seq) is a widely used method for genome-wide mapping of cDNA produced by reverse transcription that provides absolute quantification of differentially processed RNA species. When applied to cDNA produced from nascent, RNA polymerase-associated RNA, RNA-seq provides a map of the most recent output of RNA polymerases. Levels of nascent RNA can then be used to infer the effects of chromatin architecture on transcriptional output at the level of both polymerase kinetics and overall gene expression. For example, this approach was used to assess the effects of topoisomerase inhibition on gene expression in *D. melanogaster* cells[87]. As this method does not rely on affinity purification, nascent RNA-seq can in theory be applied to any RNA polymerase. However, quantification of nascent RNA species is complicated by their variable lengths from gene to gene and by co-transcriptional splicing. Nascent RNA-seq also relies on reverse transcription of RNA to cDNA prior to library construction, which may lead to biases. For example, random hexamer priming is associated with compositional biases in cDNA libraries[91], and variability in reverse transcription efficiency between transcripts due to RNA sequence and secondary structure may further distort transcript levels[92]. To circumvent these issues, direct methods have been introduced that do not rely on reverse transcription but rather determine the precise 3′ end of the nascent transcripts. In theory, such methods can provide an absolute measurement of transcription, namely the position of the base added to the nascent chain at the active site of the RNA polymerase. Three such methods (see below) accomplish precise mapping of transcription and in doing so provide a comprehensive 'snapshot' of RNA polymerases in a population of cells (FIG. 4).

NET-seq. In native elongating transcript sequencing (NET-seq)[93], cells are lysed and nascent RNA is co-purified with Pol II by ChIP. The nascent RNA is then extracted and sequenced, and the 3′ ends of sequencing reads are taken to be positions of individual Pol II molecules in the population of cells. Thus, NET-seq offers nucleotide-resolution mapping of the last base added to each transcript by Pol II and also assays both elongating and stalled or backtracked Pol II. However, the insolubility of the Pol II complex might limit its use (see above and TABLE 1).

GRO-seq and PRO-seq. Nuclear run-on is a classic assay used to measure transcription rates in isolated nuclei[94]. Nuclear run-on was adapted to the genome-wide scale as global run-on sequencing (GRO-seq)[95]. Precision nuclear run-on and sequencing (PRO-seq)[96] is a modification of GRO-seq that uses biotinylated nucleotides, which are thought to block elongation upon incorporation, to map the position of the last base added to the growing RNA chain. Moreover, these run-on-based
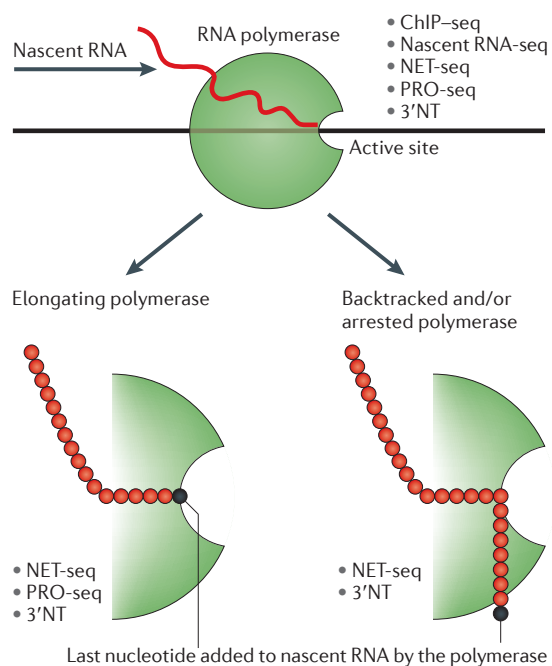

Figure 4 | Profiling transcription. An RNA polymerase molecule associated with chromatin is depicted schematically. Methods that profile the locations of polymerase on chromatin include chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq), high-throughput RNA sequencing (RNA-seq), native elongating transcript sequencing (NET-seq), precision nuclear run-on and sequencing (PRO-seq) and the 3′ nascent transcript (3′NT) method. Zoomed-in view shows the polymerase active site, with the position of the last nucleotide added to the nascent RNA chain in black in an elongating polymerase and in a backtracked and/or an arrested polymerase. Methods that can detect each form of polymerase are listed.

methods can theoretically be used to assess transcription by any RNA polymerase, as there are no polymerase-specific steps. GRO-seq and PRO-seq both require extensive nuclear manipulation and map only elongation-competent RNA polymerases; thus, they do not provide information regarding backtracked or arrested Pol II[90] (TABLE 1).

3′NT method. To comprehensively profile Pol II, a classic protocol for nascent chain isolation[97] was combined with the NET-seq library preparation protocol for mapping the 3′ base[98]. In this 3′ nascent transcript (3′NT) method, successive release of the cytoplasm and nucleoplasm and washing of the pellet result in a chromatin fraction that contains insoluble Pol II with its associated nascent transcripts. Nascent RNA is then isolated and sequenced to identify the 3′ end. This simple method can delineate the precise positions of Pol II stalling in populations of cells. The 3′NT method quantitatively maps the base incorporated at the active site of all RNA polymerases — whether they are elongating, stalled and/or backtracked — without the need for tags or affinity reagents (TABLE 1).

Nuclear run-on
A technique in which transcription is reinitiated in isolated nuclei to determine the rates at which genes are transcribed.

Although the 3′NT method was originally used to map and quantify nucleosome barriers to Pol II elongation in the *D. melanogaster* S2 cell genome and to investigate a role for the histone variant H2A.Z in modulating nucleosome barriers[98], we expect that the basic method can be applied generally to problems in transcription.

Cell-type-specific epigenome profiling

Epigenomic methods provide only an average over all cell types present in a sample. Although pure cell populations are easily obtained using cultured cells, the different cell types within an organism need to be separated to provide accurate epigenomic maps and not only averages over heterogeneous populations. We discuss below techniques that can be used to enrich specific cell populations for subsequent epigenomic analysis.

Single-cell analysis. Perhaps the best way to circumvent issues of cellular heterogeneity in epigenomic profiling is to analyse individual cells. Although single-cell genomes and transcriptomes have been generated[99], single-cell epigenomes remain elusive. Single-cell fluorescence-based methods can provide information about chromatin interaction dynamics but are extremely limited in spatial resolution. ATAC-seq, which can be used to generate epigenomic profiles using as few as 500 cells[44], may eventually prove to be suitable for single-cell epigenomic analyses.

FACS. Fluorescence-activated cell sorting (FACS)[100] uses fluorescent labelling of cells to isolate populations of interest. In this method, a fluorescent reporter is placed under the control of a cell-type-specific promoter; alternatively, fluorophore-conjugated antibodies could be directed against a specific cell surface protein. A population of cells containing the fluorophore-labelled population is then applied to a flow cytometer, in which cells are exposed to a laser emitting at the proper frequency to excite the fluorophore used. Fluorophore emission above a given threshold then signals that cell to be separated from the non-fluorescent cells. FACS isolates fairly pure cell populations but requires extensive tissue manipulation, specialized and expensive equipment, and is relatively low throughput compared with other cell separation methods.

LCM. Laser capture microdissection (LCM)[101] is a microscopy-based technique that can be used to isolate specific cell populations from tissue sections. A thermoplastic film is affixed to a region of interest selected microscopically, and a low-power infrared laser beam is directed at the cells of interest, transiently melting the film to surround the cells, which are retained within the film after its removal from the tissue section. Varying the width of the laser beam used allows isolation of both single cells and groups of cells. LCM facilitates isolation of highly pure cell populations but has similar limitations to FACS with regard to tissue manipulation and equipment. LCM is also laborious, as cells of interest must be microscopically identified, and small amounts of material are obtained with each dissection.

INTACT. Isolation of nuclei tagged in specific cell types (INTACT) uses affinity purification of nuclei to purify material for cell-type-specific epigenome analysis[102]. A nuclear envelope protein carrying an epitope tag is expressed under the control of a cell-type-specific promoter. Nuclei from the cell type of interest can then be purified on magnetic beads coated with an affinity reagent such as a tag-specific antibody. This method was used to isolate nuclei from the hair cells from the *Arabidopsis thaliana* root[102] and muscle nuclei from *C. elegans*[103] for genome-wide expression and chromatin profiling. A similar system was independently developed and applied to epigenomic and transcriptomic profiling of nuclei from *D. melanogaster* neurons[104]. INTACT requires less extensive tissue manipulations than FACS or LCM and can be carried out on the bench top with standard laboratory equipment[102].

CAST-ChIP. Chromatin affinity purification from specific cell types by chromatin immunoprecipitation (CAST-ChIP) is a variation of conventional ChIP in which a cell-type-specific promoter drives expression of an epitope-tagged general chromatin protein[105]. CAST-ChIP was used to show that the histone variant H2A.Z preferentially localized around cell-type-invariant promoters in *D. melanogaster*. As no cell or nuclear purification is required, CAST-ChIP can be applied to relatively small amounts of material for the study of chromatin regulation during development.

Perspectives

By providing high-resolution information about the structure and dynamics of the epigenome, digital epigenomic methods have the potential to provide detailed information about the mechanisms by which the myriad cell types within an organism derive their individual identities from a single genome sequence.

One of the most challenging issues facing the epigenomics field is computation. Numerous software packages are available for epigenomic data analysis, and many groups often carry out such analysis using custom software, leading to a bewildering array of options for data analysis, with different approaches giving different results in many cases. Several consortia have put forth experimental and analytical guidelines for epigenomic experiments[106], but progress in technology development makes definition of best practices a moving target, and it is likely that data analysis protocols will continue to evolve for the foreseeable future. Beyond bioinformatic analysis, the sheer volume of data generated by epigenomic methods is itself a formidable computational challenge. With decreases in sequencing cost far outpacing increases in computing power (see DNA sequencing costs), high-performance computing infrastructures are required to analyse sequencing data. Aside from processing power, storage is a major issue[107]. Raw data from a single sequencing run generally consumes several gigabytes of storage, and analytical pipelines often generate large temporary files. To publicly share epigenomic data and ease local storage constraints, raw sequencing data are often uploaded into public repositories, such as the SRA.

Despite the massive storage infrastructure in place for the SRA, the volume of sequencing data being submitted to the SRA continues to exceed the rate at which storage capacity increases[108]. The implementation of a national bioinformatic infrastructure in Sweden may hold lessons for improving processing and storage capabilities[107].

A full understanding of the epigenome requires analysis of spatiotemporal changes. Genome architecture and transcriptional activity are intimately linked[109], and DNA replication — a process known to be influenced by the epigenome[110] — is associated with changes in genome architecture[111]. Chromosome conformation capture (3C)-based methods have been extensively used to probe genome architecture. Although the resolution of 3C-based methods is still orders of magnitude less than that obtained using the digital epigenomic methods discussed here, data sets generated through epigenomic and 3C-based methods are beginning to be integrated[112,113]. By integrating three-dimensional epigenomic maps with epigenomic dynamics data, our understanding of the spatiotemporally dynamic epigenome (that is, the four-dimensional epigenome) will be enhanced.

Recent advances in genome editing technology offer great promise for facilitating epigenomics. For example, transcription activator-like effectors (TALEs) have been used to target a histone demethylase–deacetylase complex to remove histone modifications from putative enhancers, which allows possible roles of these marks in transcriptional regulation to be assessed at these specific sites[114]. In another study, TALEs were used in combination with a light-inducible system to target histone-binding proteins to endogenous loci and repress transcription[115]. Although targeting enzymes using TALEs is labour-intensive, the recent development of the CRISPR–Cas9 system holds promise to simplify genome editing for such strategies[116]. The potential for the use of CRISPR-based genome editing in conjunction with digital epigenomics is not limited to testing the effects of targeted epigenome editing. Integration of epitope tags at endogenous loci for ChIP approaches may offer a solution for the genome-wide localization of proteins for which suitable antibodies are unavailable or difficult to generate. Strategies for conditional depletion of proteins, such as the auxin degron and anchor-away systems[117,118], could be implemented using the CRISPR–Cas system and then used to assess the epigenomic effects of depleting specific factors. We look forward to the routine integration of genetic manipulation with digital epigenomics readouts to address unresolved problems in cell biology and development.

1. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nature Rev. Genet.* **10**, 161–172 (2009).
2. Talbert, P. B. & Henikoff, S. Histone variants — ancient wrap artists of the epigenome. *Nature Rev. Mol. Cell Biol.* **11**, 264–275 (2010).
3. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nature Struct. Mol. Biol.* **20**, 259–266 (2013).
4. Henikoff, S. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nature Rev. Genet.* **9**, 15–26 (2008).
5. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Rev. Genet.* **14**, 204–220 (2013).
6. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
7. Hargreaves, D. C. & Crabtree, G. R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* **21**, 396–420 (2011).
8. Flynn, R. A. & Chang, H. Y. Active chromatin and noncoding RNAs: an intimate relationship. *Curr. Opin. Genet. Dev.* **22**, 172–178 (2012).
9. Stamatoyannopoulos, J. A. What does our genome encode? *Genome Res.* **22**, 1602–1611 (2012).
10. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotech.* **28**, 1045–1048 (2010).
11. Li, R. *et al. De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).
12. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
13. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
14. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein–DNA interactions. *Science* **316**, 1497–1502 (2007).
15. Krueger, F., Kreck, B., Franke, A. & Andrews, S. R. DNA methylome analysis using short bisulfite sequencing data. *Nature Meth.* **9**, 145–151 (2012).
16. 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).
17. Reeves, R. & Jones, A. Genomic transcriptional activity and the structure of chromatin. *Nature* **260**, 495–500 (1976).
18. Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).
   References 17 and 18 describe the first uses of MNase and DNase, respectively, to map the accessibility of chromatin at specific genomic loci.
19. Sabo, P. J. *et al.* Genome-scale mapping of DNase I sensitivity *in vivo* using tiling DNA microarrays. *Nature Meth.* **3**, 511–518 (2006).
20. Crawford, G. E. *et al.* DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature Meth.* **3**, 503–509 (2006).
21. Yuan, G.-C., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. Genome-scale identification of nucleosome positions in *S. cerevisiae. Science* **309**, 626–630 (2005).
22. 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).
23. Auerbach, R. K. *et al.* Mapping accessible chromatin regions using Sono-seq. *Proc. Natl Acad. Sci. USA* **106**, 14926–14931 (2009).
24. Sulkowski, E. & Laskowski, M. Mechanism of action of micrococcal nuclease on deoxyribonucleic acid. *J. Biol. Chem.* **237**, 2620–2625 (1962).
25. Noll, M. Subunit structure of chromatin. *Nature* **251**, 249–251 (1974).
26. Weber, C. M., Henikoff, J. G. & Henikoff, S. H2A.Z nucleosomes enriched over active genes are homotypic. *Nature Struct. Mol. Biol.* **17**, 1500–1507 (2010).
27. Teves, S. S. & Henikoff, S. Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Genes Dev.* **25**, 2387–2397 (2011).
28. Kent, N. A., Adams, S., Moorhouse, A. & Paszkiewicz, K. Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. *Nucleic Acids Res.* **39**, e26 (2011).
29. Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M. & Henikoff, S. Epigenome characterization at single base-pair resolution. *Proc. Natl Acad. Sci. USA* **108**, 18318–18323 (2011).
   References 28 and 29 show that mapping of a broad range of MNase-digested fragments gives precise information about positioning and occupancy of both nucleosomes and non-histone proteins in a single sample.
30. Krassovsky, K., Henikoff, J. G. & Henikoff, S. Tripartite organization of centromeric chromatin in budding yeast. *Proc. Natl Acad. Sci. USA* **109**, 243–248 (2012).
31. Chung, H.-R. *et al.* The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS ONE* **5**, e15754 (2010).
32. Deniz, O. *et al.* Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast. *BMC Genomics* **12**, 489 (2011).
33. Allan, J., Fraser, R. M., Owen-Hughes, T. & Keszenman-Pereyra, D. Micrococcal nuclease does not substantially bias nucleosome mapping. *J. Mol. Biol.* **417**, 152–164 (2012).
34. Albert, I. *et al.* Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446**, 572–576 (2007).
35. Hesselberth, J. R. *et al.* Global mapping of protein–DNA interactions *in vivo* by digital genomic footprinting. *Nature Meth* **6**, 283–289 (2009).
36. Vierstra, J., Wang, H., John, S., Sandstrom, R. & Stamatoyannopoulos, J. A. Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nature Meth* **11**, 66–72 (2014).
37. He, H. H. *et al.* Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nature Meth* **11**, 73–78 (2014).
38. Boyle, A. P. *et al.* High-resolution genome-wide *in vivo* footprinting of diverse transcription factors in human cells. *Genome Res.* **21**, 456–464 (2011).
39. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
40. Lazarides, E. & Lindberg, U. Actin is the naturally occurring inhibitor of deoxyribonuclease I. *Proc. Natl Acad. Sci. USA* **71**, 4742–4746 (1974).
41. Grontved, L. *et al.* Rapid genome-scale mapping of chromatin accessibility in tissue. *Epigenetics Chromatin* **5**, 10 (2012).
42. 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).
43. Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S. J. & Craig, N. L. DNA transposon Hermes inserts into DNA in nucleosome-free regions *in vivo. Proc. Natl Acad. Sci.* **107**, 21966–21972 (2010).
   This study describes a rapid, simple procedure for epigenomic analysis based on transposition of sequencing adapters into chromatin.

44. 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. *Nature Meth.* **10**, 1213–1218 (2013).

45. Henikoff, S. *et al.* The budding yeast Centromere DNA Element II wraps a stable Cse4 hemisome in either orientation *in vivo*. *eLife* **3**, e01861 (2014).

46. Pan, C. Q., Landgraf, R. & Sigman, D. S. *Drosophila* engrailed-1, 10-phenanthroline chimeras as probes of homeodomain-DNA complexes. *Protein Sci.* **4**, 2279–2288 (1995).

47. Pan, C. Q., Johnson, R. C. & Sigman, D. S. Identification of new fis binding sites by DNA scission with fis-1,10-phenanthroline-copper(I) chimeras. *Biochemistry* **35**, 4326–4333 (1996).

48. Landgraf, R., Pan, C., Sutton, C., Pearson, L. & Sigman, D. S. Engineering of DNA binding proteins into site-specific cutters: reactivity of Trp repressor-1,10-phenanthroline chimeras. *Protein Eng.* **9**, 603–610 (1996).

49. Izzo, A. *et al.* The genomic landscape of the somatic linker histone subtypes H1.1 to H1.5 in human cells. *Cell Rep.* **3**, 2142–2154 (2013).

50. van Bemmel, J. G. *et al.* A network model of the molecular organization of chromatin in *Drosophila*. *Mol. Cell* **49**, 759–771 (2013).

51. Jessen, W. J. *et al.* Mapping chromatin structure *in vivo* using DNA methyltransferases. *Methods* **33**, 68–80 (2004).

52. Kelly, T. K. *et al.* Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).

53. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).

54. The modENCODE Consortium. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).

55. Wang, J. *et al.* Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res.* **41**, D171–D176 (2013).

56. Schmidt, D. *et al.* ChIP–seq: using high-throughput sequencing to discover protein–DNA interactions. *Methods* **48**, 240–248 (2009).

57. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein–DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).

58. Nakahashi, H. *et al.* A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* **3**, 1678–1689 (2013).

59. Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T. & Pugh, B. F. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149**, 1461–1473 (2012).

60. Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014). **This study suggests that TFs find their binding sites through trial-and-error sampling of degenerate motifs, which provides a potential explanation for the prevalence of low-occupancy ChIP–seq peaks with weak motifs.**

61. Guo, Y., Mahony, S. & Gifford, D. K. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput. Biol.* **8**, e1002638 (2012).

62. Serandour, A., Brown, G., Cohen, J. & Carroll, J. Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. *Genome Biol.* **14**, R147 (2013).

63. Skene, P. J., Hernandez, A. E., Groudine, M. & Henikoff, S. The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. *eLife* **3**, e02042 (2014).

64. Zentner, G. E., Tsukiyama, T. & Henikoff, S. ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. *PLOS Genet.* **9**, e1003317 (2013).

65. Kasinathan, S., Orsi, G. A., Zentner, G. E., Ahmad, K. & Henikoff, S. High-resolution mapping of transcription factor binding sites on native chromatin. *Nature Meth* **11**, 203–209 (2014).

66. Zentner, G. E. & Henikoff, S. Mot1 redistributes TBP from TATA-containing to TATA-less promoters. *Mol. Cell. Biol.* **33**, 4996–5004 (2013).

67. Orsi, G. A. *et al.* High-resolution mapping defines the cooperative architecture of Polycomb response elements. *Genome Res.* **24**, 809–820 (2014). **References 64–67 show that native ChIP is applicable to a wide range of non-histone proteins.**

68. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Rev. Genet.* **15**, 69–81 (2014).

69. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**, 1191–1200 (2002).

70. Mito, Y., Henikoff, J. G. & Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. *Nature Genet.* **37**, 1090–1097 (2005).

71. Mito, Y., Henikoff, J. G. & Henikoff, S. Histone replacement marks the boundaries of *cis*-regulatory domains. *Science* **315**, 1408–1411 (2007).

72. Chow, C.-M. *et al.* Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep.* **6**, 354–360 (2005).

73. Jin, C. *et al.* H3.3/H2A. Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature Genet.* **41**, 941–945 (2009).

74. Ooi, S. L., Henikoff, J. G. & Henikoff, S. A native chromatin purification system for epigenomic profiling in *Caenorhabditis elegans*. *Nucleic Acids Res.* **38**, e26 (2010).

75. Dion, M. F. *et al.* Dynamics of replication-independent histone turnover in budding yeast. *Science* **315**, 1405–1408 (2007).

76. Jamai, A., Imoberdorf, R. M. & Strubin, M. Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. *Mol. Cell* **25**, 345–355 (2007).

77. van Werven, F. J., van Teeffelen, H. A. A. M., Holstege, F. C. P. & Timmers, H. T. M. Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome. *Nature Struct. Mol. Biol.* **16**, 1043–1048 (2009).

78. Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**, 251–255 (2012).

79. Kraushaar, D. *et al.* Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol.* **14**, R121 (2013).

80. Huang, C. *et al.* H3.3–H4 tetramer splitting events feature cell-type specific enhancers. *PLoS Genet.* **9**, e1003558 (2013).

81. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).

82. Yang, F., Kemp, Christopher, J. & Henikoff, S. Doxorubicin enhances nucleosome turnover around promoters. *Curr. Biol.* **23**, 782–787 (2013).

83. Gupta, P., Zlatanova, J. & Tomschik, M. Nucleosome assembly depends on the torsion of the DNA molecule: a magnetic tweezers study. *Biophys. J.* **97**, 3150–3157 (2009).

84. Bermúdez, I., García-Martínez, J., Pérez-Ortín, J. E. & Roca, J. A method for genome-wide analysis of DNA helical tension by means of psoralen–DNA photobinding. *Nucleic Acids Res.* **38**, e182 (2010).

85. Naughton, C. *et al.* Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nature Struct. Mol. Biol.* **20**, 387–395 (2013).

86. Kouzine, F. *et al.* Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nature Struct. Mol. Biol.* **20**, 396–403 (2013).

87. Teves, S. S. & Henikoff, S. Transcription-generated torsional stress destabilizes nucleosomes. *Nature Struct. Mol. Biol.* **21**, 88–94 (2014).

88. Poorey, K. *et al.* Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science* **342**, 369–372 (2013). **This study indicates that a single long formaldehyde crosslinking time is unsuitable for inference of the relative occupancy or dynamics of a chromatin-binding factor.**

89. Kulaeva, O. I., Hsieh, F.-K., Chang, H.-W., Luse, D. S. & Studitsky, V. M. Mechanism of transcription through a nucleosome by RNA polymerase II. *Biochim. Biophys. Acta* **1829**, 76–83 (2013).

90. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Rev. Genet.* **13**, 720–731 (2012).

91. Hansen, K. D., Brenner, S. E. & Dudoit, S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* **38**, e131 (2010).

92. Ståhlberg, A., Håkansson, J., Xian, X., Semb, H. & Kubista, M. Properties of the reverse transcription reaction in mRNA quantification. *Clin. Chem.* **50**, 509–515 (2004).

93. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).

94. Gatehouse, J. & Thompson, A. in *Plant Gene Transfer and Expression Protocols* Vol. 49 Ch. 19, (ed. Jones, H.) 229–238 (Springer, 1995).

95. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA. Sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).

96. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).

97. Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* **14**, 7219–7225 (1994).

98. Weber, Christopher, M., Ramachandran, S. & Henikoff, S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol. Cell* **53**, 819–830 (2014).

99. Macaulay, I. C. & Voet, T. Single cell genomics: advances and future perspectives. *PLoS Genet.* **10**, e1004126 (2014).

100. Tomlinson, M. J., Tomlinson, S., Yang, X. B. & Kirkham, J. Cell separation: terminology and practical considerations. *J. Tissue Eng.* http://dx.doi.org/10.1177/2041731412472690 (2013).

101. Curran, S., McKay, J. A., McLeod, H. L. & Murray, G. I. Laser capture microscopy. *Mol. Pathol.* **53**, 64–68 (2000).

102. Deal, R. B. & Henikoff, S. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev. Cell* **18**, 1030–1040 (2010).

103. Steiner, F. A., Talbert, P. B., Kasinathan, S., Deal, R. B. & Henikoff, S. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res.* **22**, 766–777 (2012).

104. Henry, G. L., Davis, F. P., Picard, S. & Eddy, S. R. Cell type-specific genomics of *Drosophila* neurons. *Nucleic Acids Res.* **40**, 9691–9704 (2012).

105. Schauer, T. *et al.* CAST-ChIP maps cell-type-specific chromatin states in the *Drosophila* central nervous system. *Cell Rep.* **5**, 271–282 (2013).

106. Satterlee, J. S., Schubeler, D. & Ng, H.-H. Tackling the epigenome: challenges and opportunities for collaboration. *Nature Biotech.* **28**, 1039–1044 (2010).

107. Lampa, S., Dahlo, M., Olason, P., Hagberg, J. & Spjuth, O. Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data. *GigaScience* **2**, 9 (2013).

108. Kodama, Y., Shumway, M. & Leinonen, R. The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.* **40**, D54–D56 (2012).

109. Zhao, R., Bodnar, M. S. & Spector, D. L. Nuclear neighborhoods and gene expression. *Curr. Opin. Genet. Dev.* **19**, 172–179 (2009).

110. Unnikrishnan, A., Gafken, P. R. & Tsukiyama, T. Dynamic changes in histone acetylation regulate origins of DNA replication. *Nature Struct. Mol. Biol.* **17**, 430–437 (2010).

111. Kitamura, E., Blow, J. J. & Tanaka, T. U. Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell* **125**, 1297–1308 (2006).

112. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).

113. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).

114. Mendenhall, E. M. *et al.* Locus-specific editing of histone modifications at endogenous enhancers. *Nature Biotech.* **31**, 1133–1136 (2013).

115. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).

116. Rusk, N. CRISPRs and epigenome editing. *Nature Methods* **11**, 28 (2014).

117. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Meth* **6**, 917–922 (2009).

118. Haruki, H., Nishikawa, J. & Laemmli, U. K. The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol. Cell* **31**, 925–932 (2008).

119. Landt, S. G. *et al.* ChIP–seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).

120. Bailey, T. *et al.* Practical guidelines for the comprehensive analysis of ChIP–seq data. *PLoS Comput. Biol.* **9**, e1003326 (2013).

121. Wade, J. T., Struhl, K., Busby, S. J. W. & Grainger, D. C. Genomic analysis of protein–DNA interactions in bacteria: insights into transcription and chromosome organization. *Mol. Microbiol.* **65**, 21–26 (2007).

122. Biddie, S. C. *et al.* Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol. Cell* **43**, 145–155 (2011).

123. Biggin, Mark, D. Animal transcription networks as highly connected, quantitative continua. *Dev. Cell* **21**, 611–626 (2011).

124. Fisher, W. W. *et al.* DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl Acad. Sci. USA* **109**, 21330–21335 (2012). **This study shows that low-occupancy TF sites determined by ChIP–seq are often non-functional, which argues for cautious interpretation of such sites.**

125. Marinov, G. K., Kundaje, A., Park, P. J. & Wold, B. J. Large-scale quality analysis of published ChIP–seq data. *G3* **4**, 209–223 (2014). **This analysis suggests that a substantial minority of published ChIP–seq data sets are of poor or intermediate quality.**

126. Teytelman, L., Thurtle, D. M., Rine, J. & van Oudenaarden, A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc. Natl Acad. Sci. USA* **110**, 18602–18607 (2013).

127. Park, D., Lee, Y., Bhupindersingh, G. & Iyer, V. R. Widespread misinterpretable ChIP–seq bias in yeast. *PLoS ONE* **8**, e83506 (2013). **References 123 and 124 describe biases in X-ChIP–seq experiments that could lead to artefactual results.**

128. Fan, X. & Struhl, K. Where does Mediator bind *in vivo*? *PLoS ONE* **4**, e5029 (2009).

129. Jeronimo, C. & Robert, F. Kin28 regulates the transient association of Mediator with core promoters. *Nature Struct. Mol. Biol.* **21**, 449–455 (2014).

130. Zhang, L., Zhang, K., Prändl, R. & Schöffl, F. Detecting DNA-binding of proteins *in vivo* by UV-crosslinking and immunoprecipitation. *Biochem. Biophys. Res. Commun.* **322**, 705–711 (2004).

131. Vega, V. B., Cheung, E., Palanisamy, N. & Sung, W.-K. Inherent signals in sequencing-based chromatin-immunoprecipitation control libraries. *PLoS ONE* **4**, e5241 (2009).

132. Teytelman, L. *et al.* Impact of chromatin structures on DNA processing for genomic analyses. *PLoS ONE* **4**, e6700 (2009).

FURTHER INFORMATION

CexoR: www.bioconductor.org/packages/2.13/bioc/html/CexoR.html
DNA sequencing costs: genome.gov/sequencingcosts
MACE: dldcc-web.brc.bcm.edu/lilab/MACE/docs/html/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF