Review TRENDS in Genetics Vol.23 No.2

Chromatin structure in the genomics era

Oliver J. Rando

Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

The packaging of eukaryotic genomes into chromatin has a large influence on DNA-templated processes, such as transcription. The availability of genome sequences and 'genomics' technologies such as DNA microarrays and high-throughput sequencing had an immediate effect on the study of transcriptional regulation, by enabling researchers to identify the coregulation patterns of thousands of genes. These same resources are now being used successfully to study the structure of chromatin. Here, I review some of these new genomics approaches to understanding chromatin structure in eukaryotes.

Introduction

Eukaryotic DNA is packaged into a nucleoprotein complex known as chromatin, and this packaging has major functional consequences for most processes that involve DNA. Understanding processes such as DNA damage, transcriptional signal processing and cellular differentiation probably requires a detailed understanding of the chromatin context of the genome. The 'genomics era' has yielded great insights in other areas, with genome sequencing yielding extensive information about genomic organization and evolution, and with microarray measurements of transcript abundance considerably improving our understanding of transcriptional control and signal processing. The purpose of this article is to summarize the insights that have been gained by applying genomics approaches such as microarrays and high-throughput sequencing to chromatin structure.

Chromatin structure

The positioning and modification state of nucleosomes influence processes from transcription to DNA repair to replication timing. In addition to its roles in plastic responses to the environment, chromatin seems to be capable of carrying epigenetic information for many generations [1,2]. Similarly, many eukaryotes methylate a subset of cytosine bases in their genome, and this methylation regulates various DNA-templated processes (e.g. transposon silencing and telomere length maintenance) and provides a carrier for epigenetic information [3]. Both cytosine methylation and the histones are thought to be localized to a particular genomic location for longer time scales than typical DNA-binding proteins such as transcription factors (although counterexamples can be found) and, therefore, are often considered in a structural context.

As noted in Box 1, chromatin organization can conveniently be considered by analogy to protein folding. Here, I describe genomics approaches to characterizing chromatin at the levels of primary and secondary structure, and I summarize selected insights gained from these studies. The following related topics might be of interest: genomic localization of DNA-binding proteins such as transcription factors [4–6]; light-microscopy imaging of chromatin structure (e.g. in *Drosophila melanogaster* polytene chromosomes and mammalian tissue culture cells) [7,8]; and microarray studies of global gene expression changes in various chromatin mutants [9–11]. This article focuses on studies in which the structural aspects of genome packaging are measured.

Experimental approaches used for chromatin analysis

Most high-resolution genomic localization studies use DNA microarray technology as a read-out (Box 2), although sequencing has also been used as a read-out for certain experiments [12–15] and might become more widespread as high-throughput genome sequencing becomes cheaper and more accessible to most researchers [16]. Sequencing methods are widely understood: a sample of interest is prepared – by immunoprecipitation (IP), nuclease digestion or chemical treatment of DNA (to study cytosine methylation; discussed later) – and high-depth sequencing is used to characterize the isolated (or modified) DNA populations. To increase throughput, Roh *et al.* [14] used a method named serial analysis of gene expression (SAGE) [17], in which numerous short tags are prepared from isolated DNA and then concatenated to enable each sequencing read to cover numerous distinct isolated fragments.

DNA microarrays are used to characterize the differences between two populations of DNA or RNA. The two main variables in microarray studies are the type of microarray used and the method used for fractionating nucleotides before labeling and microarray hybridization. Early microarray studies were limited by the microarrays that were available, which often consisted of large (~1 kb) PCR products or short (~25–70 bp) oligonucleotides scattered through various coding regions. More recently, several companies have developed tiling microarrays, which have short oligonucleotides spaced at uniform density throughout large regions of the genome, enabling high-resolution localization measurements. Furthermore, many companies now offer

Corresponding author: Rando, O.J. (Oliver.Rando@umassmed.edu).
Available online 22 December 2006.

 0168-9525/$ – see front matter doi:10.1016/j.tig.2006.12.002

Box 1. The elements of chromatin structure

Chromatin is the nucleoprotein packaging of the eukaryotic genome, and its structure can be considered using analogies with protein structure. The primary structure – the 'sequence' – of chromatin consists of a 10 nm fiber, which is observed, when using electron microscopy, as 'beads on a string'. The repeating subunit of the primary structure of chromatin (the 'bead') is the nucleosome, which consists of 147 bp of DNA wrapped around an octamer of basic histone proteins. Nucleosomes at different locations with respect to the underlying primary structure of genomic DNA vary in subunit composition and covalent modification state, and combinations of these nucleosomes can be considered in some ways analogous to the amino acids in a protein. For example, acetylation of lysine residues in the histone tails removes a positive charge that might otherwise mediate internucleosomal interactions, so deacetylated nucleosomes are more likely to occupy positions within particular secondary structural elements. Similarly, proline typically does not occur in α helices in proteins. The number of functionally distinct 'flavors' of nucleosome is unknown. So how important is the difference between a nucleosome carrying H3K4me3K14ac (i.e. histone H3 with Lys4 trimethylated and Lys14 acetylated) and a nucleosome carrying H3K4me3K18ac? Broadly considered, at present, this is a major question in the field of chromatin structure.

The secondary structure of proteins denotes the structural folds that are present: for example, α helices. Analogously, the primary (beads on a string) structure of chromatin compacts *in vitro* into a condensed fiber called the 30 nm fiber. Recent work indicates that the 30 nm fiber consists of the 10 nm fiber folded in a zigzag manner [80]. The buffer dependence of 30-nm-fiber folding *in vitro* raises the possibility that there are other secondary structural elements *in vivo* that have not been detected owing to technical limitations. Chromatin also folds at intermediate scales between the 30 nm fiber and the whole-chromosome fold (i.e. the tertiary structure): for example, the ~80–120 kb looping that occurs between matrix-associated regions.

The tertiary structure of proteins denotes the folding of an entire polypeptide chain. The analogous structure for chromatin is the overall folding of a chromosome.

The quaternary structure of proteins denotes the packing of individual polypeptide chains. By analogy, that of chromatin denotes the orientation of different chromosomes with respect to one another in the nucleus and their positioning relative to markers such as the nuclear periphery.

custom-designed microarrays with relatively low 'up front' costs, thereby considerably lowering the barrier for entry to this area of research.

Considering fractionation methods, the most commonly used is chromatin IP (ChIP), in which formaldehyde-crosslinked chromatin fragments are isolated using antibodies that recognize specific proteins or specific proteins with particular modifications: for example, histone H4 acetylated on Lys8 (H4K8ac). Immunoprecipitated DNA is labeled and hybridized to the microarray (resulting in these studies being termed ChIP–chip studies), revealing regions of the genome that are associated with the protein of interest. Another common technique used to fractionate chromatin is nuclease digestion, because packaging of the genome into chromatin domains affects its accessibility to nucleases. Other fractionation methods that have been described to date rely on the variable solubility of formaldehyde-crosslinked chromatin in organic solvents [18], sucrose gradient fractionation based on the extent of compaction [19], and the deamination of cytosine but not methylcytosine by treatment of genomic DNA with bisulfite [13,15,20,21]. In this article, I describe some biological results obtained using these fractionation techniques.

Box 2. Sequencing versus microarray read-outs for genomic studies

To date, the two main assay technologies that have been used for genomic studies of chromatin are sequencing and microarrays. Both of these technologies can be used to characterize nucleotide populations after isolation, and both have advantages and disadvantages. Microarrays typically are used to compare the relative abundances of a collection of sequences from two samples. (Although single-color hybridization is also a widely used experimental strategy, researchers usually compare two different single-color hybridizations.) Therefore, microarrays provide relative measures of the abundance of a given sequence. Microarray studies are limited by the probe sequences chosen for the microarray, with typical numbers of sequences ranging from ~5000 to several million. Furthermore, although microarrays can distinguish sequences that differ at a single nucleotide, this is still technically challenging and requires custom microarrays. One advantage is that, at present, microarray studies are much cheaper than sequencing studies of a similar scale, and, after a microarray has been built, the time required for an experiment is short.

Although sequencing is expensive compared with microarray studies, it has certain advantages: (i) sequencing provides single-base resolution; even when using densely tiled microarrays, it is difficult to achieve single-base resolution; (ii) when coupled with an appropriate isolation method, sequencing can provide some of the advantages of single-molecule studies (e.g. the variability in the ends of molecules can be characterized); by contrast, microarrays measure the aggregate behavior of a population; (iii) sequencing studies require no assumptions about which sequences will be isolated; by contrast, microarrays are biased by the choice of sequences printed on the array.

Genomics approaches to chromatin primary structure

At least three types of inhomogeneity can affect the primary structure of chromatin: (i) nucleosomes can occlude underlying genome sequences, so the position of a nucleosome along the genome has regulatory consequences [22–25]; (ii) the histone octamer does not have a uniform composition, and alternative isoforms of canonical histones (e.g. the H3 variant H3.3) can be assembled into the octamer in place of the canonical subunit [26]; (iii) histones are subject to an astonishing number and variety of covalent modifications, and nucleosomes at different locations can differ in their pattern of covalent modification [27–29].

Nucleosome positioning and occupancy

In what was perhaps the first microarray study of nucleosome occupancy, differential partitioning of formaldehyde-crosslinked yeast chromatin during phenol–chloroform extraction was used [18]. DNA associated with small amounts of protein is found in the aqueous layer. In this study, intergenic regions were recovered preferentially in the aqueous phase, indicating that these were depleted of nucleosomes [a technique now referred to as formaldehyde-assisted isolation of regulatory elements (FAIRE)]. This was confirmed in two subsequent ChIP–chip studies in which crosslinked chromatin was immunoprecipitated using antibodies specific for histone H3 [30,31]. Low nucleosome occupancy at gene promoters has also been observed in *Drosophila*, from which nucleosomal DNA was isolated by avidin binding to biotinylated histones engineered to carry a recognition sequence for the bacterial biotin ligase BirA [32]. Hogan *et al.* have recently extended

the FAIRE technique to a dynamic setting: characterizing genome-wide variation in nucleosome occupancy during the cell cycle in yeast [33].

An alternative way to probe chromatin structure is through nuclease accessibility. DNase I and micrococcal nuclease (MNase) are the two most commonly used nucleases for these studies. MNase preferentially digests linker DNA, eventually leaving only mononucleosomal DNA, and DNase-I-hypersensitive sites have long been known to be enriched for promoters and enhancers. An early study using nuclease accessibility identified regions of compacted chromatin in the human genome by limited digestion with MNase or DNase I, followed by the isolation of poorly digested material on the basis of its solubility (after exposure to MNase) or its size as determined by electrophoresis (after exposure to DNase I) [34]. However, in this study, the resolution of the microarrays used precluded the comparison of promoters and transcribed regions, for example. More recently, higher-resolution identification of hypersensitivity to DNase I in human immune cells revealed the locations of thousands of putative regulatory elements [35,36].

In yeast, nucleosomal positions have been identified by hybridizing MNase-protected mononucleosomal DNA to high-resolution (20 bp) tiling microarrays [37]. These experiments revealed that the general nucleosome depletion observed at promoters in low-resolution studies corresponded to a long region strongly depleted of nucleosomes [these regions are typically referred to as nucleosome-free regions (NFRs), although some researchers prefer that the term NFR be qualified, because microarray studies do not provide absolute amounts and thus can only reveal relative depletion] found upstream of most yeast promoters, instead of a low occupancy of multiple closely spaced nucleosomes at promoters. The NFR contains most of the functional transcription-factor-binding sites and most of the evolutionarily conserved intergenic sequence. Furthermore, this study unexpectedly revealed that most nucleosomes are well positioned in yeast (i.e. they occupy the same position in most cells in the experimental population). Fascinatingly, a recent computational analysis implicates intrinsic nucleotide sequence preference in much of this order [38], although the number of nucleosomes positioned solely on the basis of sequence, and where factors such as internucleosomal interactions and energy-utilizing proteins act to position nucleosomes, is still unknown.

Histone-variant localization

The two histone variants that have been localized at genomic scales are the H2A variant H2A.Z (known as Htz1 in yeast) and the replication-independent H3 isoform H3.3. H3.3 was described as an H3 protein that can be assembled into chromatin in the absence of genomic replication [39]. In *Drosophila*, H3.3 was found associated with actively transcribed coding regions [32], consistent with previous data showing that H3.3 is assembled into chromatin during active RNA polymerase II (RNA Pol II)-mediated transcription [39]. Interestingly, the localization of H3.3 in these studies was biased towards the 5′-ends of coding regions, instead of being uniform across coding regions.

In yeast, several groups have studied the localization of the H2A variant Htz1. In several studies, Htz1 associated with promoters instead of coding regions, although it was excluded from the promoters of genes transcribed at extremely high levels [40–42]. Results of a higher-resolution (20 bp) study were consistent with this general localization to promoters, although the partial coverage of the yeast genome on the microarrays used (~4% of the total genome) prevented an appreciation of depletion of Htz1 at the rare genes that showed high transcription rates [43]. Taken together, results from the two highest-resolution studies [40,43] can be interpreted as showing that Htz1 typically is localized to the two nucleosomes surrounding the NFR, with a subset of promoters where Htz1 is present solely at the downstream (+1) nucleosome. The importance of this localization pattern is unknown at present, although, on the basis of studies indicating that Htz1 destabilizes nucleosomes [44,45], it has been proposed that the instability of the Htz1-containing nucleosomes facilitates their eviction during transcriptional activation [42].

Histone modification

For the broader field of signal transduction, the covalent modification of nucleosomes provides a unique opportunity to study the interactions between multiple covalent modifications. Many signaling proteins are subject to multiple covalent modifications; however, for soluble proteins, it is difficult to distinguish the differences between those carrying different combinations of modifications, except by mutagenesis studies. As an alternative, subpopulations of histones can be distinguished by their location relative to the underlying genome, and, because of this indexing, it is possible to suggest functions for different combinations of modifications on the basis of histone distributions throughout the genome.

Localization studies have been the most popular type of genomic study focused on histone modifications. Together, these studies have identified the genomic localization of almost 20 distinct histone acetylation and methylation states, on all four canonical histones and on variant histones, in wild-type budding yeast and in several mutant yeast strains [14,46–51]. Many of these modifications, together with a phosphorylation state and some methylation states that are absent from budding yeast, have also been localized in other organisms, ranging from fission yeast to *Drosophila* to human cells [52–62]. In this section, common themes of these disparate studies are summarized, using yeast chromatin as a basis for discussion. It is important to note that yeast are unicellular and, therefore, might not show the range of epigenetic variations present in metazoans; nonetheless, yeast do show chromatin-mediated epigenetic variation and have been an excellent model organism for studying epigenetic inheritance. Genomic studies in yeast are aided greatly by the compact genome of yeast, because the whole genome can be assayed with fewer microarray spots or less-extensive sequencing. Furthermore, studies in multicellular organisms must deal with multiple cell types that probably have distinctive chromatin structures. Finally, the yeast genome is repeat-poor, and repetitive sequences are difficult to analyze using microarrays or sequencing methods.

An important question in the field of chromatin research concerns the large number of histone modifications described: what is the function of having so many? One suggested answer is that having multiple independent modifications yields combinatorial complexity, resulting in a large variety of functionally distinct nucleosome 'flavors' [28]. To evaluate this suggestion requires knowledge of the number of distinct combinations that occur in the cell. Several of the genomic studies that have been discussed here evaluated the location of multiple modifications, enabling analysis of co-occurrence patterns. In all cases, it was found that groups of histone modifications co-occur at steady state, meaning that only a small subset of possible modification patterns occurs inside a cell [46,47,53,57]. For example, a single-nucleosome resolution study in yeast revealed that 12 different histone modifications co-occurred in roughly two groups [47]. Therefore, the consistent conclusion of all of these studies is that many modifications co-occur (at least in the population average), indicating that combinatorial complexity cannot be the sole reason for the huge number of histone modifications described.

The other main conclusion to be gleaned from localization studies is that various modifications are typically found in particular locations in the genome relative to underlying sequence elements (Figure 1). For example, promoters in yeast are characterized by an NFR surrounded by two nucleosomes carrying Htz1, and by low levels of H2BK16ac, H4K8ac and H4K16ac [47] and low levels of trimethylated H3K79 (H3K79me3) [49]. This is largely independent of transcription rate, although, as noted earlier, Htz1 is depleted at the promoters of extremely highly expressed genes. By contrast, several other acetylation states (H2AK7ac, H3K9ac, H3K14ac, H3K18ac, H4K5ac and H4K12ac) are found at the 5′-end of coding regions, and their levels correlate with the transcription rate [14,47,49]. Similarly, H3K4me3 is found at the 5′-ends of transcribed coding regions [47,49], whereas H3K36me3 is found at the center and 3′-ends of transcribed coding regions [49,51]. When considered at an appropriate resolution, these results can be extended to other organisms. For example, H3K4me3 and H3ac are found at the 5′-ends of transcribed sequences in human and mouse cells [53], and their levels correlate with active transcription in *Drosophila* [57].

Together, these results reveal a simpler steady-state picture of chromatin structure than might have been expected. Histone modifications do not seem to occur in huge numbers of distinct combinations, and they occur in stereotyped patterns over the genome: promoters, 5′ coding regions and 3′ coding regions can all be distinguished by their modification states. The general nature of these patterns leads to questions regarding the specificity of histone modifications in transcriptional control. In other words, if H3K4me3 is uniformly associated with the 5′-ends of actively transcribed genes, then why does only a small subset of genes show defective expression levels in mutants that completely lack this modification (i.e. *set1Δ* mutant yeast, which lack the H3K4 methylase)?

It is clear that the answer depends on the modification considered; let us consider H3K36 methylation as an instructive example. The pattern of methylated H3K36 over coding regions is biased towards the center and 3′-ends of genes, seemingly owing to the association of the methylase Set2 with the 'elongation form' of RNA Pol II [63–65]. Recent results indicate that the histone deacetylase Rpd3 is recruited to coding regions by methylated H3K36,

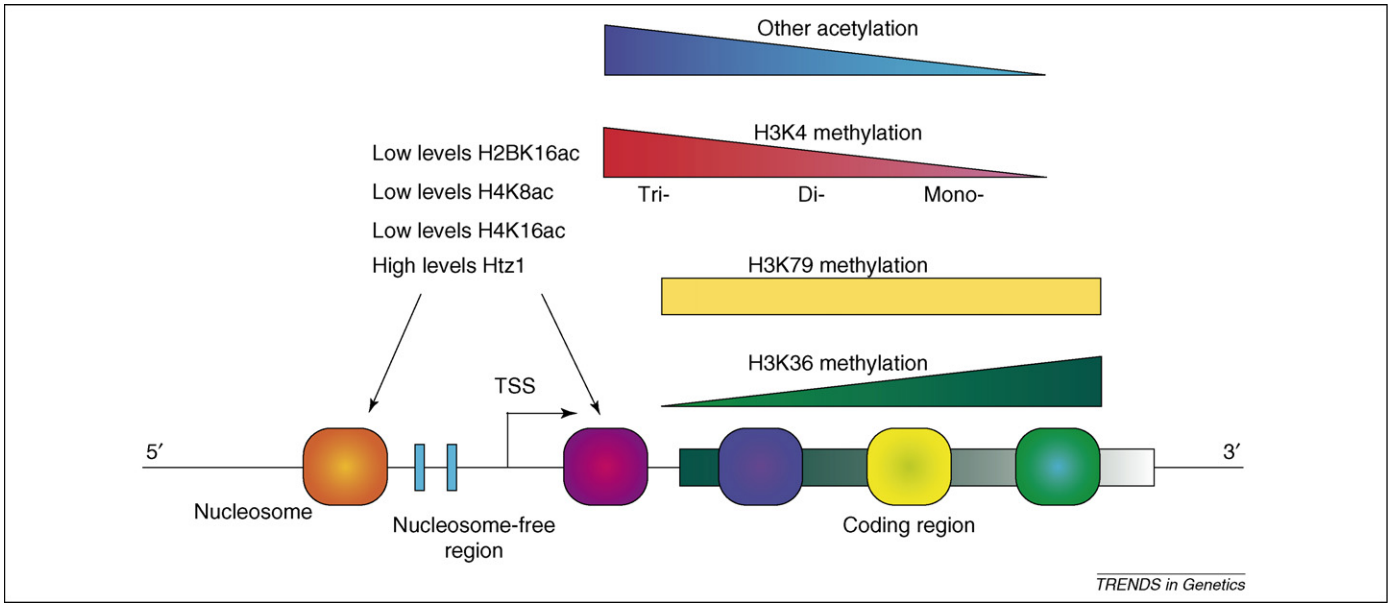

Figure 1. Nucleosome-resolution view of a typical yeast gene. Yeast genes are typically characterized by an upstream region that is nucleosome free (or at least loosely bound and highly nuclease accessible) and contains functional transcription-factor binding sites (blue rectangles) and, at its 3′-end, the transcription start site (TSS). Surrounding this nucleosome-free region are two nucleosomes that show low levels of acetylation at H2BK16, H4K8 and H4K16, and that carry Htz1 in place of the canonical histone H2A (except at high transcription rates, at which Htz1 is depleted). The remaining acetylation occurs in a gradient from the 5′-end to the 3′-end of the coding region of actively transcribed genes. Similarly, actively transcribed genes show a gradient of methylation at H3K4, with trimethylation occurring at the 5′-ends of genes and di- and monomethylation occurring in the center of the coding region. These results have been reproduced in lower-resolution studies, which have also identified that methylation at H3K79 occurs at a constant level throughout the coding region and that methylation at H3K36 occurs in a 3′-biased manner that correlates with gene expression level. Nucleosomes are colored to emphasize the different average modification patterns at each indicated location.

where it reverses the histone acetylation associated with RNA-Pol-II-mediated elongation [66–68]. In the absence of methylated H3K36, Rpd3 is not recruited, and the resultant pathological hyperacetylated state leads to transcription initiation at cryptic initiation sites within genes. Therefore, the 'purpose' of methylated H3K36 seems to be to recruit particular factors to chromatin after a round of transcription; these factors reverse the transcriptional opening of chromatin, thereby preventing inappropriate internal transcription initiation events. This example highlights a relationship between different histone modifications, the purpose of which (as it is generally understood at present) is not to generate complexity but, instead, to function as a feedback system to delimit transcriptional perturbations – coding-region histones are acetylated to enable RNA Pol II to pass through, but, after a lag, methylation prevents this open state from persisting when RNA Pol II is not actively transcribing the gene.

Genomics approaches to chromatin secondary structure

Higher-order folding of the 10 nm fiber *in vivo* has not been studied extensively at genomic scales. An important contribution to the understanding of higher-order packaging was a recent study in which chromatin from human tissue culture cells underwent limited digestion and was fractionated using a sucrose gradient [19]. At a given density level, small and large DNA fragments were compared: the expectation was that a large genomic fragment would need to be relatively compact to sediment as quickly as a short fragment, thereby enabling the identification of compacted and extended domains in the genome. This study revealed that gene-dense regions of the genome were found in extended chromatin packaging, whereas gene-poor genomic regions were compacted. Hopefully, this study will be the first of many investigations into the secondary structure of chromatin *in vivo*, which has been far less studied than the primary structure.

Beyond secondary structure, light microscopy is often used to investigate chromosome folding, but this technique seldom has single-gene resolution. A recently developed technique that holds promise for high-resolution studies of chromosomal folding is chromosome conformation capture (3C), a technique that relies on the rates of crosslinking of various genomic loci to one another to infer their physical proximity [69]. 3C was originally applied to yeast chromosome III, but it will probably be applied genome-wide to yeast and other organisms in the near future.

Genomics approaches to cytosine methylation

Most CpG dinucleotides in mammals are methylated on cytosine, and at least a portion of cytosine methylation is heritable. Cytosine methylation patterns have roles in processes ranging from development to silencing of parasitic DNA elements. In addition, as it becomes increasingly clear that inappropriate cytosine methylation of genes such as those encoding tumor suppressors can often be a key step in many cancers, interest in mapping cytosine methylation patterns has exploded [70].

A wide variety of techniques have been developed to study cytosine methylation at genomic scales. For example, antibodies specific for methylcytosine enable IP studies similar to those described earlier for examining histone modifications. Highly methylated DNA is immunoprecipitated [resulting in this technique being called mCIP, mDIP or meDIP (methylated cytosine or DNA IP)] using this antibody and is hybridized to microarrays. Studies using this approach have reported loci that are differentially methylated between samples from tumors and samples not from tumors [71,72]. However, detailed follow-up studies with single-cytosine resolution often reveal that differentially methylated loci identified by microarray studies differ in an all-or-none manner over multiple CpGs, indicating that meDIP might not be able to identify single-cytosine differences in methylation. A recent landmark study in *Arabidopsis thaliana* used much higher-resolution (35 bp) microarrays to analyze immunoprecipitated DNA. Nonetheless, the stringent thresholds used for methylated regions resulted in the elimination of sparsely methylated domains, such as the 3′-end of the *FWA* locus, where only two CpG sites are methylated [73]. Therefore, it seems unlikely that meDIP will cover most of the genome sequence at single-cytosine discrimination for methylation.

Beyond IP studies, cytosine methylation has properties that distinguish it from typical DNA-associated factors such as proteins: namely, cytosine methylation inhibits the activity of some restriction enzymes. This enables researchers to compare digestion using methylation-sensitive enzymes with digestion using enzymes that recognize the same DNA sequence but are not influenced by methylation. Many of these pairs of enzymes exist, and various studies have used single enzymes or cocktails of these enzymes [56,74–77]. This approach has single-cytosine resolution but is limited to assaying cytosines that are located in appropriate restriction sites (e.g. C**C**GG for *Hpa*II).

Finally, cytosine methylation influences the chemical reactivity of cytosine. Treating genomic DNA with bisulfite deaminates unmethylated cytosine while leaving methylcytosine intact. This difference has been exploited in two ways. First, bisulfite-treated DNA can be labeled for hybridization to oligonucleotide microarrays printed using a pair of oligonucleotides corresponding to every assayed CpG, one with CG intact (corresponding to methylated C) and one with TG in its place (corresponding to an unmodified C) [20,21]. However, both of these studies used PCR after bisulfite conversion of DNA, limiting the coverage of the assay to the number of PCR reactions that were multiplexed. In principle, this assay provides almost single-cytosine resolution (depending on CpG density and oligonucleotide length), although it is not clear how well this approach scales up, particularly given the reduced complexity of bisulfite-converted DNA. A similar approach has been used at a somewhat higher throughput for studying the human genome: bisulfite-converted genomic DNA was assayed for single-nucleotide polymorphisms (C versus T, in this case) using primer extension, followed by microarray measurement of ~1500 CpG sites [78]. This approach requires multiple oligonucleotides for each assayed CpG but has been successfully scaled to >1000 CpGs.

Second, an alternative assay for cytosine methylation would be to sequence the entire bisulfite-treated genome,

as this approach would, in principle, yield complete coverage of the genome at single-cytosine resolution and would identify all bisulfite-sensitive methylation, including any non-CpG methylation. This has been initiated [13,15], although a complete genome resequencing has not yet been reported, and given the interest in surveying various tissues in physiological and pathological states, this approach is prohibitively costly and time-consuming for most individual researchers.

Conclusions

Localization studies enabled by the genomics era are providing researchers with what can be considered to be sequencing technologies for chromatin. Already, these studies have greatly enriched our understanding of chromatin structure in the cell, providing intimate views of the steady-state patterns of chromatin and cytosine methylation in a variety of organisms and cell types. Several avenues of inquiry should prove fruitful in the future. One approach is the use of existing technologies to study cells in dynamic settings (e.g. in response to the environment and in cell-cycle-synchronized populations). This should help to tease apart colocalization patterns and to identify transient structural characteristics that are lost in the steady state. This effort is likely to be further aided by studying mutant organisms (e.g. histone acetylase mutants) in the same conditions to help determine the interdependencies (i.e. the crosstalk) between various modifications. Another approach is to develop new technologies that address the following: (i) chromatin structure in single cells [79]; (ii) structural characteristics such as DNA bulges on individual nucleosomes; (iii) dynamic properties such as the half-life of acetyl groups at various genomic locations; and (iv) locations of higher-order structural elements (e.g. the 30 nm fiber).

Acknowledgements

I thank Bing Ren for critically reviewing the manuscript before publication. I apologize to researchers whose work was not cited owing to space limitations. O.J.R. is supported in part by a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund, by the National Institute of General Medical Sciences and by the Human Frontier Science Program.

References

1 Francis, N.J. and Kingston, R.E. (2001) Mechanisms of transcriptional memory. *Nat. Rev. Mol. Cell Biol.* 2, 409–421
2 Ringrose, L. and Paro, R. (2004) Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* 38, 413–443
3 Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33, 245–254
4 Lieb, J.D. *et al.* (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association. *Nat. Genet.* 28, 327–334
5 Ren, B. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309
6 van Steensel, B. and Henikoff, S. (2000) Identification of *in vivo* DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* 18, 424–428
7 Silver, L.M. and Elgin, S.C. (1976) A method for determination of the *in situ* distribution of chromosomal proteins. *Proc. Natl. Acad. Sci. U. S. A.* 73, 423–427
8 Turner, B.M. *et al.* (1990) Islands of acetylated histone H4 in polytene chromosomes and their relationship to chromatin packaging and transcriptional activity. *J. Cell Sci.* 96, 335–346
9 Dion, M.F. *et al.* (2005) Genomic characterization reveals a simple histone H4 acetylation code. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5501–5506
10 Martin, A.M. *et al.* (2004) Redundant roles for histone H3 N-terminal lysine residues in subtelomeric gene repression in *Saccharomyces cerevisiae*. *Genetics* 167, 1123–1132
11 Parra, M.A. *et al.* (2006) Deciphering the roles of the histone H2B N-terminal domain in genome-wide transcription. *Mol. Cell. Biol.* 26, 3842–3852
12 Roh, T.Y. *et al.* (2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* 19, 542–552
13 Rakyan, V.K. *et al.* (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* 2, e405 DOI: 10.1371/journal.pbio.0020405 (biology.plosjournals.org)
14 Roh, T.Y. *et al.* (2004) High-resolution genome-wide mapping of histone modifications. *Nat. Biotechnol.* 22, 1013–1016
15 Meissner, A. *et al.* (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877
16 Shendure, J. *et al.* (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309, 1728–1732
17 Velculescu, V.E. *et al.* (1995) Serial analysis of gene expression. *Science* 270, 484–487
18 Nagy, P.L. *et al.* (2003) Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6364–6369
19 Gilbert, N. *et al.* (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* 118, 555–566
20 Gitan, R.S. *et al.* (2002) Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.* 12, 158–164
21 Scholz, C. *et al.* (2005) Distinction of acute lymphoblastic leukemia from acute myeloid leukemia through microarray-based DNA methylation analysis. *Ann. Hematol.* 84, 236–244
22 Archer, T.K. *et al.* (1991) Transcription factor access is mediated by accurately positioned nucleosomes on the mouse mammary tumor virus promoter. *Mol. Cell. Biol.* 11, 688–698
23 Lomvardas, S. and Thanos, D. (2002) Modifying gene expression programs by altering core promoter chromatin architecture. *Cell* 110, 261–271
24 Martinez-Campa, C. *et al.* (2004) Precise nucleosome positioning and the TATA box dictate requirements for the histone H4 tail and the bromodomain factor Bdf1. *Mol. Cell* 15, 69–81
25 Stunkel, W. *et al.* (1997) A nucleosome positioned in the distal promoter region activates transcription of the human U6 gene. *Mol. Cell. Biol.* 17, 4397–4405
26 Henikoff, S. and Ahmad, K. (2005) Assembly of variant histones into chromatin. *Annu. Rev. Cell Dev. Biol.* 21, 133–153
27 Schreiber, S.L. and Bernstein, B.E. (2002) Signaling network model of chromatin. *Cell* 111, 771–778
28 Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature* 403, 41–45
29 Nightingale, K.P. *et al.* (2006) Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. *Curr. Opin. Genet. Dev.* 16, 125–136
30 Bernstein, B.E. *et al.* (2004) Global nucleosome occupancy in yeast. *Genome Biol.* 5, R62 DOI: 10.1186/gb-2004-5-9-r62 (genomebiology.com)
31 Lee, C.K. *et al.* (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* 36, 900–905
32 Mito, Y. *et al.* (2005) Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* 37, 1090–1097
33 Hogan, G.J. *et al.* (2006) Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet.* 2, e158 DOI: 10.1371/journal.pgen.0020158 (genetics.plosjournals.org)
34 Weil, M.R. *et al.* (2004) Global survey of chromatin accessibility using DNA microarrays. *Genome Res.* 14, 1374–1381
35 Crawford, G.E. *et al.* (2006) DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat. Methods* 3, 503–509
36 Sabo, P.J. *et al.* (2006) Genome-scale mapping of DNase I sensitivity *in vivo* using tiling DNA microarrays. *Nat. Methods* 3, 511–518

37 Yuan, G.C. *et al.* (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309, 626–630
38 Segal, E. *et al.* (2006) A genomic code for nucleosome positioning. *Nature* 442, 772–778
39 Ahmad, K. and Henikoff, S. (2002) The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* 9, 1191–1200
40 Guillemette, B. *et al.* (2005) Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol.* 3, e384 DOI: 10.1371/journal.pbio.0030384 (biology.plosjournals.org)
41 Li, B. *et al.* (2005) Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18385–18390
42 Zhang, H. *et al.* (2005) Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* 123, 219–231
43 Raisner, R.M. *et al.* (2005) Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. *Cell* 123, 233–248
44 Abbott, D.W. *et al.* (2001) Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation. *J. Biol. Chem.* 276, 41945–41949
45 Suto, R.K. *et al.* (2000) Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat. Struct. Biol.* 7, 1121–1124
46 Kurdistani, S.K. *et al.* (2004) Mapping global histone acetylation patterns to gene expression. *Cell* 117, 721–733
47 Liu, C.L. *et al.* (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol.* 3, e328 DOI: 10.1371/journal.pbio.0030328 (biology.plosjournals.org)
48 Millar, C.B. *et al.* (2006) Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes Dev.* 20, 711–722
49 Pokholok, D.K. *et al.* (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122, 517–527
50 Robyr, D. *et al.* (2002) Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* 109, 437–446
51 Rao, B. *et al.* (2005) Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol. Cell. Biol.* 25, 9447–9459
52 ENCODE Project Consortium. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306, 636–640
53 Bernstein, B.E. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169–181
54 Bernstein, B.E. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326
55 Boyer, L.A. *et al.* (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353
56 Lippman, Z. *et al.* (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471–476
57 Schubeler, D. *et al.* (2004) The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* 18, 1263–1271
58 Sinha, I. *et al.* (2006) Genome-wide patterns of histone modifications in fission yeast. *Chromosome Res.* 14, 95–105
59 Wiren, M. *et al.* (2005) Genomewide analysis of nucleosome density histone acetylation and HDAC function in fission yeast. *EMBO J.* 24, 2906–2918
60 Tolhuis, B. *et al.* (2006) Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. *Nat. Genet.* 38, 694–699
61 Squazzo, S.L. *et al.* (2006) Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Res.* 16, 890–900
62 Schwartz, Y.B. *et al.* (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* 38, 700–705
63 Li, B. *et al.* (2003) The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* 278, 8897–8903
64 Phatnani, H.P. *et al.* (2004) Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry* 43, 15702–15719
65 Xiao, T. *et al.* (2003) Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev.* 17, 654–663
66 Carrozza, M.J. *et al.* (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581–592
67 Joshi, A.A. and Struhl, K. (2005) Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell* 20, 971–978
68 Keogh, M.C. *et al.* (2005) Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123, 593–605
69 Dekker, J. *et al.* (2002) Capturing chromosome conformation. *Science* 295, 1306–1311
70 Feinberg, A.P. *et al.* (2006) The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.* 7, 21–33
71 Keshet, I. *et al.* (2006) Evidence for an instructive mechanism of *de novo* methylation in cancer cells. *Nat. Genet.* 38, 149–153
72 Weber, M. *et al.* (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37, 853–862
73 Zhang, X. *et al.* (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126, 1189–1201
74 Yamada, Y. *et al.* (2004) A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res.* 14, 247–266
75 Tompa, R. *et al.* (2002) Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr. Biol.* 12, 65–68
76 Ching, T.T. *et al.* (2005) Epigenome analyses using BAC microarrays identify evolutionary conservation of tissue-specific methylation of SHANK3. *Nat. Genet.* 37, 645–651
77 Khulan, B. *et al.* (2006) Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* 16, 1046–1055
78 Bibikova, M. *et al.* (2006) High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* 16, 383–393
79 Fatemi, M. *et al.* (2005) Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res.* 33, e176 DOI: 10.1093/nar/gni180 (nar.oxfordjournals.org)
80 Schalch, T. *et al.* (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436, 138–141