Human Molecular Genetics, 2009, Vol. 18, Review Issue 2 ***R195–R201***
doi:10.1093/hmg/ddp409

Predictive chromatin signatures in the mammalian genome

Gary C. Hon[1,2], R. David Hawkins[2] and Bing Ren[1,2,3,*]

[1]Bioinformatics Program, [2]Ludwig Institute for Cancer Research and [3]Department of Cellular and Molecular Medicine, University of California, San Diego School of Medicine, 9500 Gilman Drive #0653, CMM East Room 3080, La Jolla, CA 92093-0653, USA

Received July 15, 2009; Revised and Accepted August 21, 2009

The DNA sequence of an organism is a blueprint of life: it harbors not only the information about proteins and other molecules produced in each cell, but also instructions on when and where such molecules are made. Chromatin, the structure of histone and DNA that has co-evolved with eukaryotic genome, also contains information that indicates the function and activity of the underlying DNA sequences. Such information exists in the form of covalent modifications to the histone proteins that comprise the nucleosome. Thanks to the development of high throughput technologies such as DNA microarrays and next generation DNA sequencing, we have begun to associate the various combinations of chromatin modification patterns with functional sequences in the human genome. Here, we review the rapid progress from descriptive observations of histone modification profiles to highly predictive models enabling use of chromatin signatures to enumerate novel functional sequences in mammalian genomes that have escaped previous detection.

INTRODUCTION

Each of the over 200 cell types in the human body contains a nearly identical copy of the genome sequence. Yet the gene expression pattern for each distinct cell type is unique (1,2). Although it is generally accepted that this uniqueness arises from differences in how transcription factors act in each cell, it has also become clear that chromatin remodeling plays an integral role in the process. Gene regulation is the result of the interactions of these factors with chromatin and a host of genomic regulatory elements including promoters, enhancers, silencers and insulators (3). Experimentally, promoters and insulators can be identified through binding of RNA polymerase II (RNAPII) and the insulator-binding protein CCCTC-binding factor (CTCF), respectively (4,5). In contrast, most other *cis*-regulatory elements have remained undiscovered, though techniques based on evolutionary conservation have identified a handful of these elements. But recently, it has been shown that chromatin modifications can be regarded as indicators of the transcriptional regulatory function and activity of certain types of genomic loci. With recent technological advances making it routine to survey chromatin modifications on a large scale, the epigenetics field is rapidly expanding from examining individual genes to all genes to the entirety of the human genome. With this extension comes the recent shift from descriptive to predictive models relating chromatin signatures and the regulatory elements they mark, giving new global insights into gene regulation and development by allowing dissection of these processes in unprecedented detail.

CHROMATIN SIGNATURES AT GENE STRUCTURES

One of the most studied chromatin modifications is histone H3 lysine 4 tri-methylation (H3K4me3), a hallmark of actively transcribed protein-coding promoters in eukaryotes spanning yeast to human (Fig. 1). This has been made clear first through chromatin immunoprecipitation (ChIP) studies focusing on individual promoters, then through small scale ChIP followed by microarray analysis (ChIP-chip) studies spanning the full yeast genome or subsets of the human genome (5–8), and most recently through genome-wide techniques using ChIP-chip or ChIP followed by sequencing (ChIP-Seq) (9–11). Newer technologies have offered higher resolution views, showing clearly that this modification is found on the nucleosomes flanking nucleosome-free regions that coincide with the transcription start sites (TSSs) of actively transcribed genes in all eukaryotes examined to date (6,12–14).

*To whom correspondence should be addressed. Tel: +1 8588225766; Fax: +1 8585347750; Email: biren@ucsd.edu

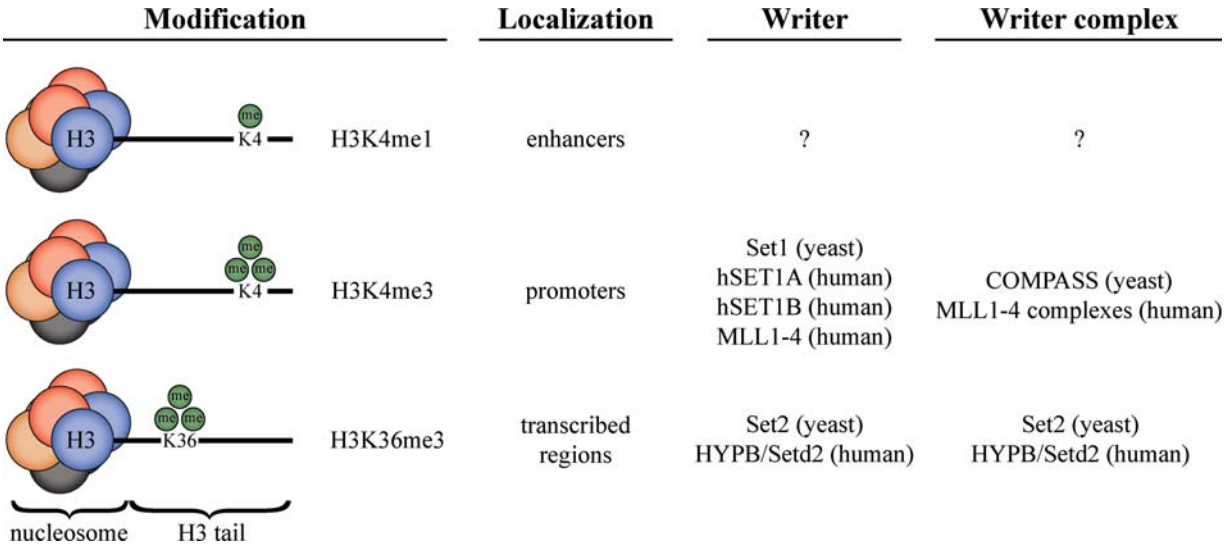

Figure 1. Modifications of histone H3. Lysine residues on histone H3 can be mono-, di- or tri-methylated. Shown are modifications H3K4me1, H3K4me3 and H3K36me3, which mark active/poised enhancers, active/poised promoters and actively transcribed regions, respectively. me, methylation.

It is now reasonably clear how the H3K4me3 chromatin signature is deposited at active promoters (Figs 1 and 2). Examining a panel of histone methyl-transferases in *Saccharomyces cerevisiae*, Briggs *et al*. (15) observed that only knock-out of the Set1 methyltransferase results in complete loss of H3K4me3. Krogan *et al*. (16,17) demonstrated that the Set1-containing COMPASS complex is recruited to active promoters by the elongating factor Paf1. Using ChIP-chip, Ng *et al*. (18) observed that Set1 occupied actively transcribed regions, and its recruitment depends on the active form of RNAPII bearing a serine-5 phosphorylated tail. More recently, Lee *et al*. (19) showed that di- and tri-methylation of H3K4 at active promoters also requires another histone modification, mono-ubiqitinylation of H2BK123, catalyzed by Rad6/Bre1 in yeast. Thus, a complex pathway that involves multiple protein complexes and crosstalks between different chromatin modifications is responsible for H3K4me3 at yeast promoters. In flies and mammals, a similar pathway has been identified that results in H3K4me3 at active promoters. At least six mammalian homologs of COMPASS exist, including MLL1–4 complexes, hSET1A and hSET1B, and their recruitment to active promoters can result in H3K4me3 (20–22).

Another canonical histone modification found in genic regions is H3K36me3, which has long been associated with the gene bodies of actively transcribed genes. Owing to the low resolution of traditional ChIP, ChIP-qPCR, and non-overlapping tiled microarrays used in ChIP-chip, this modification was long thought to be just a signal of elongation that is enriched non-specifically throughout the entire transcribed region (14). Recent observations using higher resolution techniques have instead found that enrichment of H3K36me3 is much higher at exons than introns in *Caenorhabditis elegans*, mouse, and human (23) (Figs 1 and 2). The observation that a chromatin signature marks exons has lent further support to the view that transcription and splicing are coupled events, implying that the complex processes regulating splicing may be controlled at the chromatin level (24).

Like H3K4me3, the H3K36me3 chromatin signature is coordinated by phosphorylation states of RNAP II. In *S. cerevisiae*, during transcriptional initiation, the C-terminal domain (CTD) of RNAPII is phosphorylated at serine-5, which recruits the Set1 protein to catalyze trimethylations of H3K4 (18). After promoter clearance and during transcriptional elongation, serine-5 phosphorylation of the CTD is replaced by serine-2 phosphorylation (25) and, as a result, Set1 is dissociated—explaining why H3K4me3 is not deposited in the gene body (18). Instead, serine-2 phosphorylation leads to recruitment of Set2, which results in trimethylation of H3K36me3 in the gene body (16).

Although these results show that RNAPII-transcribed sequences are generally marked by consistent chromatin signatures, an open question is whether the same chromatin signatures also exist for genes transcribed by the other polymerases. There are several polymerases known to exist in eukaryotes, each of which transcribes a distinct class of functional elements. It is possible that these distinct polymerases deposit distinct chromatin modifications during transcription. However, the chromatin signatures observed for RNAPII-transcribed genes may arise because of the highly regulated nature of RNAPII transcription. Given that RNAPI and RNAPIII generally transcribe ubiquitously expressed elements such as rRNAs and tRNAs, less regulation of this process is required, which may require fewer chromatin modifications.

TOWARDS PREDICTIVE CHROMATIN SIGNATURES

A central barrier to our understanding of the human genome is an incomplete annotation of the elements encoded in it. Many human functional elements have been assigned on the basis of sequence homology with other species under the assumption that sequence conservation equates functional conservation

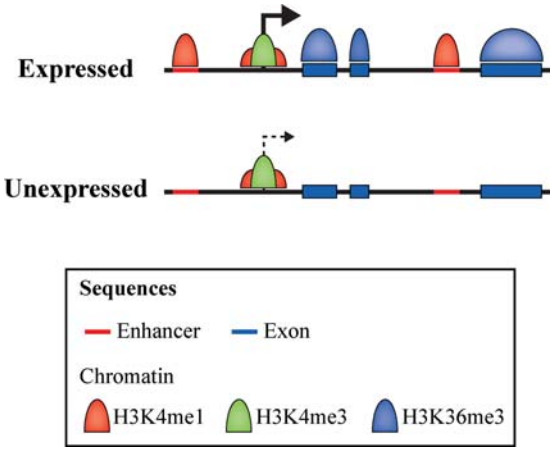

Figure 2. Chromatin signatures and gene expression. The promoters of both actively expressed and unexpressed genes are marked by H3K4me3. In contrast, only actively expressed genes have exons marked by H3K36me3 and nearby enhancers marked by H3K4me1.

(26,27). These techniques by definition miss human or lineage-specific elements, which may be important in defining human-specific traits. Desperately needed are general, cost–efficient methods to identify functional elements in the human genome using only measurements from human cells. The observations that chromatin signatures are found at well-annotated places of the genome and that their presence correlates with activity suggest that examination of the human epigenome can reveal the functional elements contained within it.

But this requires a shift from descriptive to predictive models of chromatin. The vast majority of studies profiling chromatin structure have focused on the descriptive view that functional loci contain chromatin signatures; for example, the statement that active promoters are marked by H3K4me3. A stronger statement would be that H3K4me3 only marks active promoters. This predictive view suggests that the presence of the chromatin signature alone can predict the presence of a specific class of functional element. This second view is more rigorous, offers a computational strategy to identify functional elements and outlines specifically how to test hypotheses of function.

Several studies have shown that novel promoters can be identified on the basis of the H3K4me3 chromatin signature. Work from our laboratory, as well as by other groups, have shown that this mark can be used in conjunction with others to efficiently identify promoters for known and novel protein-coding genes (6,28,29). Focusing on the 1% of the human genome studied by the ENCODE pilot project (30), we identified 198 regions bearing the promoter chromatin signature (6). While the vast majority of these are recovered by known annotations, six were novel. Using luciferase reporter assays, we verified that several of these novel chromatin signature-based promoter predictions showed promoter activity *in vivo*. Similarly, taking advantage of the observations that miRNAs are transcribed by the same machinery as protein-coding genes and have promoters marked by nucleosome-depleted TSSs flanked by H3K4me3, Ozsolak *et al*. (31) were able to precisely map the locations of miRNA TSSs in human.

The most exciting applications of predictive chromatin signatures are in the identification of previously elusive regulatory elements. For example, isolated examples of non-protein-coding RNAs (ncRNAs) such as HOTAIR, which regulates expression of HOX cluster genes (32), have suggested a crucial role of ncRNAs in development. However, studies of ncRNAs are hindered by small catalogs of known ncRNA genes. To address this problem, Guttman *et al*. (33) took advantage of the observation that many non-protein-coding genes are also transcribed by the same machinery as coding genes, with RNAPII as the central component. Since RNAPII-associated enzymes deposit H3K4me3 at promoters during initiation and H3K36me3 during elongation to mark the direction of transcription, Guttman *et al*. (33) searched for this chromatin signature in several mouse strains. This approach successfully identified over 1000 ncRNAs including well-known members such as HOTAIR. Subsequent analysis revealed that these ncRNAs show complex expression and regulatory patterns similar to those previously observed for protein-coding genes, suggesting they are functional during mouse development.

PREDICTIVE CHROMATIN SIGNATURES AT ENHANCERS

Chromatin modifications associated with genes and gene-proximal elements have been extensively investigated, either through closed experimental systems that exclusively survey genic regions (14,34,35) or through exclusively analyzing only these regions even when using open experimental systems that survey the entire genome (9,10,36). But, perhaps owing to their ambiguous link to gene expression, chromatin modifications outside of genic regions has remained largely unexplored, even though epigenetic events outside of genes likely contribute to controlling gene expression.

In eukaryotes, transcription is tightly regulated by the activity of transcription factors, many of which bind to enhancers that are far from the genes they activate (3). As such, identifying active enhancers on a genome-wide scale has been an open problem. Our laboratory had previously shown that active transcriptional enhancers are marked by a distinct and predictive chromatin signature, central to which is strong enrichment of H3K4me1 (6) (Figs 1 and 2). Recently, we have used this well-defined chromatin signature to map 55 000 enhancers genome-wide in several human cell lines (29). Unlike promoters and insulators, the chromatin modifications marking enhancers are highly cell-type specific in a manner that correlates with cell-type specific gene expression. These results tie the global architecture of chromatin signatures outside genes to regulation of gene expression.

Unlike H3K4me3 and H3K36me3, H3K4me1 is relatively less well studied. It is still unclear what enzyme is responsible for depositing H3K4me1 at transcriptional enhancers. It is possible that this mark arises from *de novo* addition of a methyl group to an unmodified H3K4 residue or via demethylation from di or tri-methylated states. The latter would require an intermediate state containing either H3K4me2 or H3K4me3 but that also shows no promoter activity. It is possible that these intermediate states are short-lived and hard to detect by high throughput studies in large population of cells.

Although we have observed a handful of promoter-distal hypersensitive loci marked with stronger enrichment of H3K4me2 than H3K4me1 that may be places being demethylated to the mono-methylated form, this evidence is anecdotal at best and does not convincingly demonstrate the phenomenon on a large scale (Hon *et al*., unpublished data).

It will be intriguing to learn what large protein complex, if any, is responsible for depositing H3K4me1. This complex may be a key regulator of enhancer activity. Although H3K4me3 and H3K36me3 are intimately linked to phosphorylation states of CTD in RNAPII, it is unlikely that RNAPII plays a similar role with maintenance of H3K4me1 given that the observation that enhancers are generally not enriched for RNAPII (6). This factor or set of factors must satisfy several conditions. First, given that enhancers are typically found far from gene regions, the factor's binding to DNA cannot be limited to gene regions. Second, being a general factor required for enhancer activity, it must be ubiquitously expressed in all cell types. Thirdly, the factor should be capable of binding inactive, unmarked enhancers. Pioneer factors such as FoxA1, which can bind repressed enhancers enveloped by heterochromatin and open them for activity, or proteins associated with pioneer factors, would satisfy these constraints (37,38).

Thus far, the only epigenetic modification predictive of active enhancers is H3K4me1. Finding other predictive modifications or modifiers of enhancer activity has been an active area of research. Computationally, Won *et al*. (28) have employed hidden markov models (HMMs) and simulated annealing to identify histone modifications that are optimal predictors of both active promoters and enhancers. Experimentally, using a technique called GMAT that involves ChIP followed by serial analysis of gene expression-like sequencing, Roh *et al*. (39) identified thousands of acetylation islands marked by H3K9ac or H3K14ac in human cells. But since H3K9ac is known to mark the activity of promoters more than enhancers (29), the majority of these acetylation islands were close to TSSs. Although promoter-specific acetylations have been discovered (40), thus far there have been no reports of acetylations specific to enhancers. Instead, acetylation of histones has generally been associated with active chromatin regions marking both promoters and enhancers (29,40).

Using ChIP-Seq to map a panel of histone modifications (10), Barski *et al*. observed that enhancers were marked by H3K4me3. This apparent enrichment of H3K4me3 at enhancers could be caused by secondary physical interactions between H3K4me3-marked promoters and H3K4me1-marked enhancers as predicted by the looping mechanism of enhancer activity (41). Indeed, analysis of this data reveals that H3K4me1 enrichment is much stronger than H3K4me3 at enhancers. H3K4me3 enrichment at enhancers could also be attributed to certain technical limitations of ChIP-Seq, which is inherently biased towards identifying highly fragile genomic regions, a feature that characterize both promoters and enhancers (42). In contrast, in ChIP-chip studies where the ChIP sample is hybridized together with a genomic control sample, H3K4me3 is rarely observed above background levels at enhancers. Normalization procedures that take into account of input control may relieve the observed H3K4me3 enrichment.

SYSTEMATIC DISCOVERY OF CHROMATIN SIGNATURES

Increasingly, we are coming to appreciate the chromatin epigenome as a cell-type specific interpretation of the genetic code, specifying the activity of every part of the genome. The observations that chromatin signatures are predictive of a variety of transcribed elements including promoters, exons, miRNAs and ncRNAs as well as untranscribed regulatory elements such as enhancers leads one to suspect that novel chromatin signatures may also mark other elements of unique function.

Nucleosome depletion is common among many active regulatory elements including genic promoters and miRNA promoters. We have also observed that enhancers marked by H3K4me1 are depleted for core histone H3 (6,29). Like the distribution of nucleosomes around an active TSS, nucleosomes may be well-positioned flanking a region of nucleosome depletion around other active genomic regions. Indeed, we observe a bimodal distribution of H3K4me1 enrichment at predicted human enhancers, and most interestingly enrichment of transcription factors is strongest inside the nucleosome-free region. Similar observations have been observed at CTCF-bound insulators (43). Recently developed technologies such as DNase-Seq are enabling the efficient enumeration of all nucleosome-free regions in the human genome (44), and systematic examination of the histone modifications around these regions will likely yield novel chromatin signatures of enhancers and other regulatory elements.

Not all functional regions of the genome are expected to be marked by DNase I hypersensitivity, particularly places of the genome that are repressed. One way to find consistent chromatin signatures marking regulatory elements outside of known regulatory regions or annotations is to apply an unbiased search on multiple dimensions of the chromatin epigenome simultaneously. This can be achieved by using various machine learning techniques. We have recently developed a computational technique called ChromaSig to identify frequently occurring chromatin signatures (45) (Fig. 3). Focusing on genomic regions with strong enrichment of histone modifications, ChromaSig employs a probabilistic approach to simultaneously align and cluster these regions to identify consistent signatures. In agreement with observations in yeast (46), we find that many histone modifications are highly redundant, resulting in only a handful of distinct chromatin signatures in the human genome (45). In particular, we observe classes of inactive genomic regions marked by multiple, distinct repressive chromatin modifications but that are not DNase I hypersensitive (45). More recently, HMMs have also been extended to genome-wide histone modification profiles with the goal of partitioning the entire genome into contiguous blocks, each with distinct combinations of histone modifications (47,48) (Ernst and Kellis, unpublished data). In so doing, these initial studies are expanding our catalogs of chromatin signatures, linking them with functional elements, and detailing how chromatin states transition into others on a chromosomal level.

Chromatin structure is constantly changing in response to the cell's many stimuli. In addition to defining how gene expression is presently controlled, chromatin modifications also detail how the cell is ready to respond to environmental or developmental cues to alter its transcriptional output. This

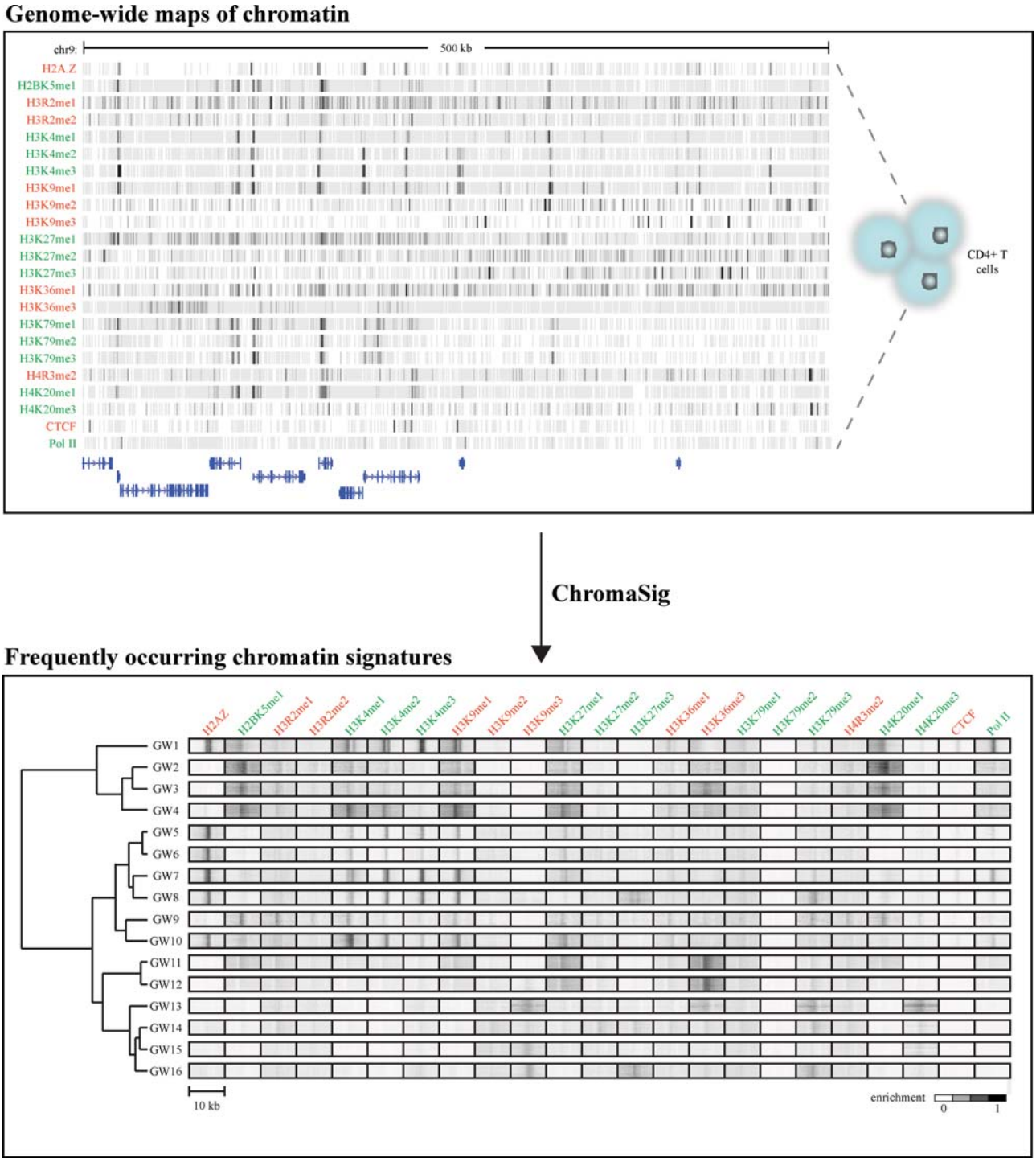

Figure 3. Identifying chromatin signatures with ChromaSig. (Top) A snapshot of histone methylations, CTCF and RNA polymerase II mapped using ChIP-Seq in CD4+ T cells (10). White bars indicate no enrichment, whereas black bars indicate enrichment. (Bottom) Frequently occurring chromatin signatures found by ChromaSig in examining genome-wide maps of histone methylations.

poised phenomenon has been well documented at promoters where a bivalent chromatin state ensures a poised transcriptional state critical for development (49), and likely also applies to enhancers (6,29,37) and, by extension, to other regulatory elements. Using unbiased approaches to identify which parts of the epigenome change during cellular response will reveal key regulatory elements involved in the process. Most interesting will be identifying which parts of the genome are marked both before and after stimulation, but where the marks have significantly changed either in terms of modification types or spatial distribution. These poised elements may be those most critical in defining the cellular response.

CONCLUSIONS

To dissect the human genome, we must first enumerate all the regulatory elements encoded by it. Although we know that many classes of functional elements exist, current approaches to map these elements are not general, efficient, accurate, genome-scale and cell-type specific. A major obstacle in finding these elements from the genome sequence alone is that there are no natural breaks in the sequence that delimit phrases or functional elements. The epigenome is an interpretation of the genome. But although the alphabet of the chromatin epigenome is larger than that of the genome, its analysis is a much more tractable endeavor as the words of histone modification peaks are well spaced throughout the genome. Furthermore, as the fundamental unit of this chromatin epigenome is the nucleosome, it is effectively orders of magnitude shorter than the genome, telling the story of the genome in a more compact way without skipping the important features. Well defined, predictive chromatin signatures offer an elegant framework to comprehensively map all the functional elements in the human genome.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by funding from American Cancer Society (R.D.H.), LICR (B.R., G.C.H.), NHGRI (B.R.), NIEHS (B.R.) and CIRM (B.R.).

REFERENCES

1. Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A. *et al.* (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl Acad. Sci. USA*, **99**, 4465–4470.
2. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
3. Maston, G.A., Evans, S.K. and Green, M.R. (2006) Transcriptional Regulatory Elements in the Human Genome. *Annu. Rev. Genomics Hum. Genet.*, **7**, 29–59.
4. Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V. and Ren, B. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, **128**, 1231–1245.
5. Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D. and Ren, B. (2005) A high-resolution map of active promoters in the human genome. *Nature*, **436**, 876–880.
6. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A. *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, **39**, 311–318.
7. Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P. *et al.* (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.*, **17**, 691–707.
8. Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M. *et al.* (2005) Direct isolation and identification of promoters in the human genome. *Genome Res.*, **15**, 830–839.
9. Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553–560.
10. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
11. Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R. and Thomson, J.A. (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell*, **1**, 299–312.
12. Ozsolak, F., Song, J.S., Liu, X.S. and Fisher, D.E. (2007) High-throughput mapping of the chromatin structure of human promoters. *Nat. Biotechnol.*, **25**, 244–248.
13. Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G. and Zhao, K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell*, **132**, 887–898.
14. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. and Young, R.A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, **130**, 77–88.
15. Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F. and Allis, C.D. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.*, **15**, 3286–3295.
16. Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C. *et al.* (2003) Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.*, **23**, 4207–4218.
17. Krogan, N.J., Dover, J., Khorrami, S., Greenblatt, J.F., Schneider, J., Johnston, M. and Shilatifard, A. (2002) COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J. Biol. Chem.*, **277**, 10753–10755.
18. Ng, H.H., Robert, F., Young, R.A. and Struhl, K. (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell*, **11**, 709–719.
19. Lee, J.S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R. and Shilatifard, A. (2007) Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell*, **131**, 1084–1096.
20. Kim, J., Guermah, M., McGinty, R.K., Lee, J.S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W. and Roeder, R.G. (2009) RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell*, **137**, 459–471.
21. Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H. and Allis, C.D. (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, **121**, 859–872.
22. Shilatifard, A. (2008) Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr. Opin. Cell Biol.*, **20**, 341–348.
23. Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S. and Ahringer, J. (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.*, **41**, 376–381.
24. Sims, R.J. III and Reinberg, D. (2009) Processing the H3K36me3 signature. *Nat. Genet.*, **41**, 270–271.
25. Komarnitsky, P., Cho, E.J. and Buratowski, S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.*, **14**, 2452–2460.
26. Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
27. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
28. Won, K.J., Chepelev, I., Ren, B. and Wang, W. (2008) Prediction of regulatory elements in mammalian genomes using chromatin signatures. *BMC Bioinformatics*, **9**, 547.
29. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, **459**, 108–112.
30. ENCODE_Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636–640.

31. Ozsolak, F., Poling, L.L., Wang, Z., Liu, H., Liu, X.S., Roeder, R.G., Zhang, X., Song, J.S. and Fisher, D.E. (2008) Chromatin structure analyses identify miRNA promoters. *Genes Dev.*, **22**, 3172–3183.
32. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323.
33. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
34. Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
35. Miao, F. and Natarajan, R. (2005) Mapping global histone methylation patterns in the coding regions of human genes. *Mol. Cell. Biol.*, **25**, 4650–4661.
36. Cui, K., Zang, C., Roh, T.Y., Schones, D.E., Childs, R.W., Peng, W. and Zhao, K. (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell*, **4**, 80–93.
37. Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S. and Brown, M. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*, **132**, 958–970.
38. Eeckhoute, J., Lupien, M., Meyer, C.A., Verzi, M.P., Shivdasani, R.A., Liu, X.S. and Brown, M. (2009) Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res.*, **19**, 372–380.
39. Roh, T.Y., Cuddapah, S., Cui, K. and Zhao, K. (2006) The genomic landscape of histone modifications in human T cells. *Proc. Natl Acad. Sci. USA*, **103**, 15782–15787.
40. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q. *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, **40**, 897–903.
41. Bulger, M. and Groudine, M. (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.*, **13**, 2465–2477.
42. Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M. and Gerstein, M.B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat. Biotechnol.*, **27**, 66–75.
43. Fu, Y., Sinha, M., Peterson, C.L. and Weng, Z. (2008) The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.*, **4**, e1000138.
44. Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S. and Crawford, G.E. (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell*, **132**, 311–322.
45. Hon, G., Ren, B. and Wang, W. (2008) ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput. Biol.*, **4**, e1000201.
46. Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N. and Rando, O.J. (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol.*, **3**, e328.
47. Jaschek, R. and Tanay, A. (2009) Spatial clustering of multivariate genomic and epigenomic information. *Lecture Notes in Computer Science*, **5541**, 170–183.
48. Jia, L., Landan, G., Pomerantz, M., Jaschek, R., Herman, P., Reich, D., Yan, C., Khalid, O., Kantoff, P., Oh, W. *et al.* (2009) Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet.*, **5**, e1000597.
49. Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 12.