

# Mapping Human Epigenomes

Chloe M. Rivera<sup>1,2</sup> and Bing Ren<sup>1,3,\*</sup>

<sup>1</sup>Ludwig Institute for Cancer Research

<sup>2</sup>The Biomedical Sciences Graduate Program

<sup>3</sup>Department of Cellular and Molecular Medicine

Institute of Genomic Medicine, UCSD Moores Cancer Center, University of California School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093-0653, USA

\*Correspondence: [biren@ucsd.edu](mailto:biren@ucsd.edu)

<http://dx.doi.org/10.1016/j.cell.2013.09.011>

As the second dimension to the genome, the epigenome contains key information specific to every type of cells. Thousands of human epigenome maps have been produced in recent years thanks to rapid development of high throughput epigenome mapping technologies. In this review, we discuss the current epigenome mapping toolkit and utilities of epigenome maps. We focus particularly on mapping of DNA methylation, chromatin modification state, and chromatin structures, and emphasize the use of epigenome maps to delineate human gene regulatory sequences and developmental programs. We also provide a perspective on the progress of the epigenomics field and challenges ahead.

## Introduction

More than a decade has passed since the human genome was completely sequenced, but how genomic information directs spatial- and temporal-specific gene expression programs remains to be elucidated (Lander, 2011). The answer to this question is not only essential for understanding the mechanisms of human development, but also key to studying the phenotypic variations among human populations and the etiology of many human diseases. However, a major challenge remains: each of the more than 200 different cell types in the human body contains an identical copy of the genome but expresses a distinct set of genes. How does a genome guide a limited set of genes to be expressed at different levels in distinct cell types?

Overwhelming evidence now indicates that the epigenome serves to instruct the unique gene expression program in each cell type together with its genome. The word “epigenetics,” coined half a century ago by combining “epigenesis” and “genetics,” describes the mechanisms of cell fate commitment and lineage specification during animal development (Holliday, 1990; Waddington, 1959). Today, the “epigenome” is generally used to describe the global, comprehensive view of sequence-independent processes that modulate gene expression patterns in a cell and has been liberally applied in reference to the collection of DNA methylation state and covalent modification of histone proteins along the genome (Bernstein et al., 2007; Bonasio et al., 2010). The epigenome can differ from cell type to cell type, and in each cell it regulates gene expression in a number of ways—by organizing the nuclear architecture of the chromosomes, restricting or facilitating transcription factor access to DNA, and preserving a memory of past transcriptional activities. Thus, the epigenome represents a second dimension of the genomic sequence and is pivotal for maintaining cell-type-specific gene expression patterns.

Not long ago, there were many points of trepidation about the value and utility of mapping epigenomes in human cells

(Madhani et al., 2008). At the time, it was suggested that histone modifications simply reflect activities of transcription factors (TFs), so cataloging their patterns would offer little new information. However, some investigators believed in the value of epigenome maps and advocated for concerted efforts to produce such resources (Feinberg, 2007; Henikoff et al., 2008; Jones and Martienssen, 2005). The last five years have shown that epigenome maps can greatly facilitate the identification of potential functional sequences and thereby annotation of the human genome. Now, we appreciate the utility of epigenomic maps in the delineation of thousands of lincRNA genes and hundreds of thousands of *cis*-regulatory elements (ENCODE Project Consortium et al., 2012; Ernst et al., 2011; Guttman et al., 2009; Heintzman et al., 2009; Xie et al., 2013b; Zhu et al., 2013), all of which were obtained without prior knowledge of cell-type-specific master transcriptional regulators. Interestingly, bioinformatic analysis of tissue-specific *cis*-regulatory elements has actually uncovered novel TFs regulating specific cellular states.

Propelled by rapid technological advances, the field of epigenomics is enjoying unprecedented growth with no sign of deceleration. An expanding cadre of researchers is working to explore exciting frontiers in epigenomics. Many international consortia have been formed to tackle the fundamental problems in epigenomics by sharing resources and protocols (Table 1) (Beck et al., 2012; Bernstein et al., 2010). Consequently, the number of epigenomic data sets and publications has grown exponentially in recent years. The resulting epigenomic maps have linked genomic sequences to many nuclear processes including splicing, replication, DNA damage response, folding, chromatin packaging, and cell-type-specific gene expression patterns.

In this review, we focus on recent progress in several areas of epigenomics (Table 2). First, we describe the remarkable advances in epigenomic technologies, especially next-generation sequencing-based applications, which have fueled the

**Table 1. Large-Scale National and International Epigenomic Consortia**

Project Name	Start Date	Affiliations	Completed and Expected Data Contributions	Selected Publication	Access Data
Encyclopedia of DNA Elements	2003	NIH	Dnase-seq, RNA-seq, ChIP-seq, and 5C in 100s of primary human tissues and cell lines	<a href="#">ENCODE Project Consortium et al., 2012</a>	<a href="http://encodeproject.org/ENCODE/">http://encodeproject.org/ENCODE/</a>
The Cancer Genome Atlas (TCGA)	2006	NIH	DNA methylomes in 1,000s of patients samples from more than 20 cancer types	<a href="#">Garraway and Lander, 2013</a>	<a href="http://cancergenome.nih.gov/">http://cancergenome.nih.gov/</a>
Roadmap Epigenomics Project	2008	NIH	Dnase-seq, RNA-seq, ChIP-seq, and MethylC-seq in 100 s of normal primary cells, hESC, and hESC derived cells	<a href="#">Bernstein et al., 2010</a>	<a href="http://www.epigenomebrowser.org/">http://www.epigenomebrowser.org/</a>
International Cancer Genome Consortium (ICGC)	2008	15 countries, includes TCGA	DNA methylation profiles in thousands of patient samples from 50 different cancers	<a href="#">The International Cancer Genome Consortium, et al., 2010</a>	<a href="http://dcc.icgc.org/web">http://dcc.icgc.org/web</a>
International Human Epigenome Consortium (IHEC)	2010	7 countries, includes BLUEPRINT, Roadmap	Goal: 1,000 Epigenomes in 250 cell types	<a href="#">American Association for Cancer Research Human Epigenome Task Force; European Union, Network of Excellence, Scientific Advisory Board, 2008</a>	<a href="http://ihc-epigenomes.org">http://ihc-epigenomes.org</a>

growth of the field. Second, we discuss the utility of epigenomic maps, emphasizing the power of these maps in annotating transcription units and *cis*-regulatory elements in the context of development and disease pathogenesis. We also explore new biological insights gained through integrative analysis of epigenomic maps in mammalian cell systems, highlighting the study of pluripotency and lineage specification of embryonic stem cells. Finally, we provide a perspective on the road ahead regarding meeting the technical challenges and addressing unanswered questions in the field. As for advancements in our understanding of the mechanistic roles of sequence-specific TFs and noncoding RNAs, chromatin, DNA modifying enzymes, and chromatin-binding proteins in establishing, maintaining, and removing epigenetic marks, we refer readers to recent excellent reviews ([Badeaux and Shi, 2013](#); [Calo and Wysocka, 2013](#); [Lee and Young, 2013](#); [Pastor et al., 2013](#); [Rinn and Chang, 2012](#); [Smith and Meissner, 2013](#)).

### Epigenome Mapping Technologies

The unprecedented genome-wide scope and nucleotide precision with which we can now map human epigenomes was enabled by disruptive technology advancement, namely the application of microarrays and next-generation sequencing (NGS). A slew of molecular biology assays previously used to measure a single locus are now integrated with these platforms. Powerful parallel short-read sequencing technologies have proven increasingly high-throughput, fast, accurate, and cost-effective at rates faster than Moore's Law. By far, the greatest advantage of NGS is its ability to survey the entire genome in an unbiased and comprehensive manner. Due to this monumental shift in assay capacity, researchers can ask if conclusions drawn from locus-centered studies extend to other

parts of the genome. In defiance of the traditional scientific method emphasizing hypothesis testing, global profiling promotes hypothesis-free exploration of new observations and correlations. Overall, the accomplishments and discoveries of the epigenomics field have hinged in large part on the iterative inventions and improvements of technologies ([Figure 1](#)). Here, we discuss cutting-edge epigenome mapping techniques that integrate NGS platforms and disclose their advantages and disadvantages.

### Mapping DNA Methylation

Of the four nucleotides composing DNA, cytosine is by far the most dynamic. Cytosine can be methylated at its 5<sup>th</sup> carbon (5mC) and in the human genome 60%–80% of the 28 million CpG dinucleotides are methylated ([Lister et al., 2009](#); [Ziller et al., 2013](#)). Stressing the importance of DNA methylation during development, deletion of cytosine methyltransferases responsible for de novo (DNMT3A, DNMT3B) or maintenance (DNMT1) of methylation through cellular divisions results in embryonic and neonatal lethality in mice ([Li et al., 1992](#); [Okano et al., 1999](#)). Detection of DNA methylation at individual loci and with promoter-focused studies established the important repressive roles of DNA methylation in imprinting, retrotransposon silencing, and X chromosome inactivation ([Bird, 2002](#)). Global DNA methylation technologies now measure DNA methylation abundance at all cytosines at base resolution in the human genome. The elucidation of complete human methylomes progressed the narrow view that 5mC is only a stable repressive mark to the view that it is an epigenetic mark that is dynamically deposited and removed, can exist in non-CpG sequence contexts, and is enriched at the bodies of actively transcribed genes ([Hellman and Chess, 2007](#); [Lister et al., 2009](#)).

**Table 2. Major Conceptual Advances Convey the Utility of Epigenome Maps**

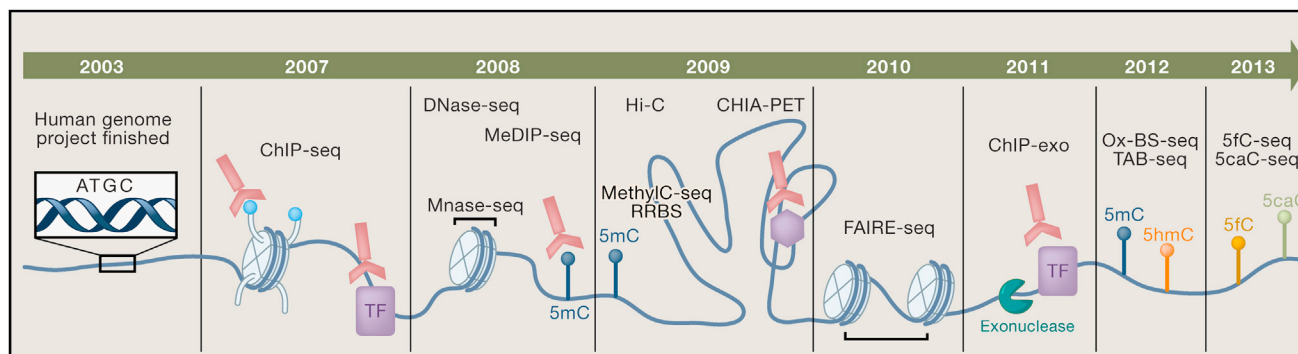
Before Next-Gen Sequencing	The Next-Gen Sequencing Era	Future <sup>a</sup>
<b>DNA Methylation</b>		
Repressive mark at imprinted loci, transposons, and in x chromosome inactivation	Metastable mark	Comprehensive definition of epigenomic variation across all human cell types
Found in active gene bodies	Active DNA demethylation by TET family proteins through 5hmC, 5fC, and 5caC	Define epigenomic variation in populations
Only dynamic in primordial germ cells and during early embryogenesis; in CpG context	Non-CpG methylation exists especially in ESCs oocytes, and adult brain	Discovering epigenetic signatures of disease
	Tissue-specific methylation at distal regulatory elements	Technology advances to enable single-cell methylomes
<b>Histone Modifications</b>		A deeper understanding of epigenomic inheritability and reprogramming
Marks that correlate with promoters and gene bodies	Over 130 different histone modifications have been identified	High-throughput functional validation of predicted enhancers
H3K9me3 is a mark of heterochromatin	Identification of novel ncRNAs by promoter and gene body chromatin signatures	Epigenome engineering
H3K27me3 is a mark of facultative heterochromatin	Active enhancers are marked by H3K27ac or H4K16ac	Improved computational analysis and visualization tools
Proposal of the Histone Code Hypothesis	Poised enhancers are marked by H3K4me1 alone or in combination with H3K27me3	
Bivalent Promoters marked by H3K4me3/H3K27me3	Expansion of repressive chromatin blocks during differentiation	
Unique chromatin signature of enhancers defined as H3K4me1	Combinations of chromatin marks define a limited number of chromatin states	
<b>Chromatin Structure</b>		
Nucleosome maps only in yeast, fly	Nucleosomes mapped in the human genome	
DHSs correlate with TF-binding sites and regulatory elements	Nucleosomes are well positioned around regulatory regions	
	DHSs predict enhancer-promoter pairs and cell types affected in disease	
<b>Nuclear Architecture</b>		
Identification of chromosome territories, LADS, and transcription factories with FISH	Identification of sub-TADs, TADs and chromosome compartments	
Few validated enhancer-promoter interacting pairs	Chromosome-wide maps of enhancer, promoter, and insulator interacting pairs	

<sup>a</sup>Text in this column refers to DNA Methylation, Histone Modifications, Chromatin Structure, and Nuclear Architecture.

The DNA methylation toolkit includes three main molecular-biology-based techniques: digestion of genomic DNA with methyl-sensitive restriction enzymes, affinity-based enrichment of methylated DNA fragments, and chemical conversion methods (Bock, 2012; Laird, 2010). The choice of how to assay DNA methylation depends on the resolution and genome coverage needs, and both parameters ultimately dictate the experimental cost (Figure 2). Endonuclease digestion-based DNA methylation assays (MRE-seq, etc.) determine CpG methylation at base resolution by digesting genomic DNA with several methylation-sensitive enzymes. The genome-wide CpG coverage of MRE-seq is limited by the cutting frequency of the chosen restriction enzyme(s) and biased toward the enzyme recognition sequences, whereas the accuracy is dependent on complete digestion. Affinity-based enrichment assays capture methylated fragments from sonicated DNA with an antibody (MeDIP-seq) or a methyl-binding domain (MBD-seq) (Down

et al., 2008; Serre et al., 2010). When sequencing enriched DNA fragments, at least one cytosine is certainly methylated but the exact site or combination of sites can not be directly determined. Therefore, the resolution of affinity-based assays is highly dependent on the DNA fragment size, CpG density, and immunoprecipitation quality of the reagent. The results of both restriction-enzyme- and affinity-based sequencing methods are qualitative levels of enrichment rather than absolute. On the other hand, because affinity- and restriction-enzyme-based methylation assays enrich or capture methylated DNA regions the sequencing costs are moderate.

Bisulfite sequencing is a chemical conversion method that directly determines the methylation state of each cytosine in a binary fashion and is widely accepted as a gold standard for mapping DNA methylation. Treatment of genomic DNA with sodium bisulfite chemically converts unmethylated cytosines to uracil. After PCR, and assuming nearly complete bisulfite



**Figure 1. Timeline of Sequencing-Based Technologies for Mapping Human Epigenomes**

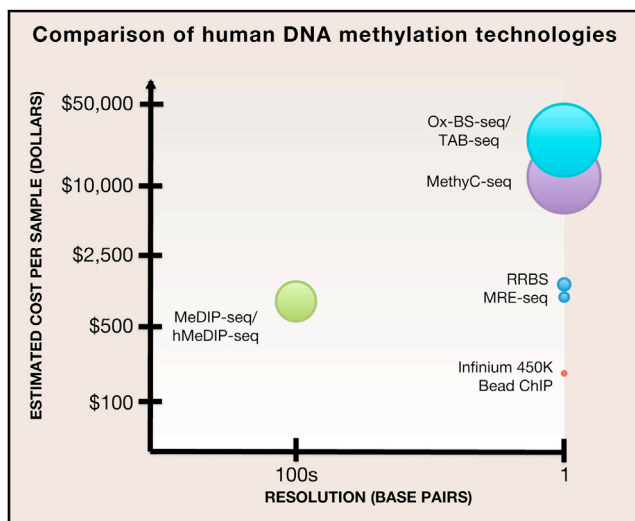
conversion, all unmethylated cytosines become thymidines and remaining cytosines correspond to 5mC. Initially, individual loci were assayed from BS-treated genomic DNA with locus-specific PCR followed by Sanger sequencing (Clark et al., 1994). In a step toward increasing genomic coverage, reduced representation bisulfite sequencing (RRBS) combines restriction digestion with bisulfite sequencing for specific interrogation of high CpG density regions such as clusters of CpGs at promoters called CpG islands (Meissner et al., 2005). At the pinnacle of the genomic coverage spectrum, bisulfite treatment coupled with whole-genome sequencing (variably referred to as MethylC-seq, BS-seq, or WGBS) features nucleotide resolution and quantitative rates of methylation for all cytosines (Cokus et al., 2008; Lister et al., 2008). Careful analysis of MethylC-seq data should consider PCR biases due to unbalanced GC content of methylated and unmethylated fragments, mapping inefficiencies of bisulfite-treated DNA, and estimate the bisulfite conversion rate using a spike-in control (Krueger et al., 2012; Laird, 2010).

The DNA methylation field collectively experienced an epiphany upon the discovery of 5-hydroxymethylcytosine (5hmC) as an intermediate of demethylation of 5mC to cytosine (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). The ten-eleven translocation (TET) family of proteins, TET1, TET2, and TET3, oxidize 5mC through 5hmC, 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) intermediates before being replaced by cytosine via base excision repair pathways or a yet-unidentified decarboxylase (Pastor et al., 2013). As it became clear that four variants of cytosine exist, a clarification to MethylC-seq results also came to light. 5mC and 5hmC, but not 5fC or 5caC, are both resistant to bisulfite conversion and therefore cannot be distinguished from each other in MethylC-seq data (Huang et al., 2010; Jin et al., 2010). In order to understand the role of DNA demethylation, new techniques would need to be developed to accurately differentiate cytosine methylation states.

In an exciting advancement in the field, an assortment of methods was published for the detection of all cytosine methylation states across the genome. The first versions use antibodies to either directly immunoprecipitate 5hmC (hMeDIP-seq) or chemically modify 5hmC making it more immunogenic (anti-CMS, hMe-Seal, GLIB, and JBP-1) (Ficz et al., 2011; Pastor et al., 2011; Robertson et al., 2011; Song et al., 2011; Williams et al., 2011). Although these methods are an important advance-

ment, affinity-based techniques are hindered by low-resolution and qualitative signal, as mentioned above. Last year, both oxidative bisulfite sequencing (oxBS-seq) and TET-assisted bisulfite sequencing (TAB-seq) were introduced as single-base resolution methods for measuring 5hmC (Booth et al., 2012; Yu et al., 2012). In oxBS-seq, 5hmC nucleotides are sensitized to bisulfite treatment after a chemical reaction specifically oxidizes all 5hmC and to 5fC. Then, DNA is successively treated with sodium bisulfite such that remaining cytosines must originally be 5mC. The sequencing results of oxBS-seq, therefore, accurately capture 5mC levels and subtraction from MethylC-seq data reveals true 5hmC sites. Of concern with oxBS-seq are the successive chemical treatments that can induce DNA damage and skew the results. The second approach, TAB-seq, directly assays 5hmC location and abundances. First, 5hmC is tagged with a glucose molecule using the T4 bacteriophage enzyme beta glucosyltransferase. Next, genomic DNA is treated with purified TET enzyme to oxidize 5mC to 5caC while glucosylated-5hmC is protected. Finally, after bisulfite treatment and sequencing, only 5hmC is read as cytosine, while all unmethylated cytosines and cytosine variants are detected as thymidine. TAB-seq's accuracy is especially dependent on efficient oxidation and conversion by TET, such that a bottleneck is the tedious process of purifying catalytically active TET enzyme (Song et al., 2012). In a final round of progress for DNA methylation detection, methods to detect 5fC (fC-seq, fCAB-seq, 5fC-DP-Seq) and 5caC (5caC-seq) were recently described (Raiber et al., 2012; Shen et al., 2013; Song et al., 2013).

Methods without enrichment steps, like MethylC-seq, oxBS-seq, and TAB-seq, require an immense amount of sequencing and are costly (Figure 2). In order to quantitatively measure the methylation rate at each cytosine, high-quality single-nucleotide resolution 5mC methylome data sets are expected to have 30X sequencing depth. A recent study by Meissner and colleagues finds that only ~20% of CpGs are differentially methylated between 30 diverse human cell and tissue types tested. Therefore, up to 80% of MethylC-seq data are the result of superfluous sequencing of DNA fragments without CpGs or containing uninformative, constitutively methylated CpGs across 30 cell and tissue types examined (Ziller et al., 2013). Sequencing costs can be limited by targeted DNA methylation mapping as



**Figure 2. Comparison of Human DNA Methylation Mapping Technologies**

The size of the circle is proportional to the genome-wide CpG coverage of the technology.

accomplished by bisulfite padlock probes (BSPPs) and Illumina's Infinium 450K BeadChIP technology (Bibikova et al., 2011; Diep et al., 2012). Theoretically, base resolution methylomes of rare cytosine variants such as 5caC should have 1,000X sequence coverage (Pastor et al., 2013). The extreme sequencing depth is necessitated by the rarity of cytosine variants; for example, in mouse embryonic stem cells (mESCs) 5hmC accounts for about 0.1% and 5caC accounts for 0.0003% of cytosines (Ito et al., 2011; Yu et al., 2012). Unfortunately, the prohibitive cost of genome-wide base resolution human methylomes has limited the number of available data sets, and presumably slowed progress toward novel insights. Resolution, genomic coverage, and monetary cost serve as interdependent considerations for generating methylome datum, where one parameter cannot be changed without affecting the others (Figure 2). Maximizing resolution and coverage while keeping costs low, currently an unrealistic situation, may eventually be possible with innovative sequencing platforms or a yet undiscovered alternative to bisulfite treatment. We look forward to the democratization of nucleotide resolution DNA methylation technologies and the resulting novel findings from a diverse collection of methylomes.

### Mapping Chromatin Modification States

Chromosomal DNA is packaged into nucleosomes with DNA wrapped around histone octamers consisting of H2A, H2B, H3, H4 subunits and their variants. The histone tails and globular domains of histone proteins are subject to over 130 posttranslational modifications (PTMs) and over 700 distinct histone isoforms have been detected in human cells (Tan et al., 2011; Tian et al., 2012). Well-studied covalent modifications on histones include methylation, acetylation, phosphorylation, and ubiquitination. State-of-the-art mass-spectrometry-based proteomic technologies unveiled many novel histone PTMs, such

as crotonylation, succinylation, malonylation, and others (Tian et al., 2012). Histone modifications serve both activating and silencing roles in transcription, generally by controlling the accessibility of DNA and by serving as binding substrates that recruit or exclude protein complexes (Kouzarides, 2007). For example, H3K27ac is found at both active promoters and enhancers, H3K36me3 identifies actively transcribed gene bodies, and H3K27me3 marks heterochromatic or repressed regions (Li et al., 2007) (Table 3). Determining the genome-wide distribution of a histone mark can lead to clues about its role in transcriptional regulation and provoke follow-up mechanistic studies to further understand the PTMs deposition, removal, and role in development and disease.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) maps the genome-wide binding pattern of chromatin-associated proteins, which includes modified histones. To perform this method, DNA-protein complexes containing a specific protein of interest are immunoprecipitated from cross-linked, sonicated chromatin. DNA is purified from the enriched pool and adaptors are ligated for subsequent PCR and sequencing. The digital sequences of enriched DNA, called reads, are computationally aligned to the reference genome to define punctate peaks or broad blocks of modified histones or protein occupancy. Since its development in 2007, researchers have used ChIP-seq extensively to survey the genomic profiles of histones and their modifications, TFs, DNA and histone modifying enzymes, transcriptional machinery, and other chromatin-associated proteins (Barski et al., 2007; Johnson et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007). Furthermore, multidimensional data sets are now available for cell lines, primary cells, tissues, and embryos from an increasing number of species. The ENCODE and Roadmap Epigenomics Projects have contributed enormous data repositories by performing thousands of ChIP-seq experiments in hundreds of human cell types.

As the number of ChIP-seq data sets began to grow exponentially, lab-to-lab protocol variability threatened the quality of results and downstream cross-study analysis. To stave off data inconsistencies, both the ENCODE and Roadmap consortia published optimized standard operating procedures for ChIP-seq. Of major concern was the quality of antibodies for which ChIP is undeniably dependent. Both consortia assessed histone modification antibody quality using dot blot immunoassays against histone tail peptides to ensure specific binding and minimal cross reactivity (Egelhofer et al., 2011). The ENCODE Project's best practices include a rigorous two-stage antibody validation process with a combination of immunoassays, immunofluorescence patterns, and functional assays (Landt et al., 2012). These screening methods dramatically improved the data quality and reduced costs and lost time from failed ChIP-seq experiments. However, the membrane-binding conditions in immunoassays are not identical to immunoprecipitation in solution using magnetic beads. Lamentably, the gold standard for ChIP-grade antibody classification remains actually performing the assay itself. To this point, Bernstein and colleagues designed a method called ChIP-string to screen for effective antibodies against chromatin regulator proteins (Ram et al., 2011). In this approach, multiplexed meso-scale ChIP-seq experiments



**Table 3. Distinctive Chromatin Features of Genomic Elements**

Functional Annotation	Histone Marks	References
Promoters	H3K4me3	Bernstein et al., 2005; Kim et al., 2005; Pokholok et al., 2005
Bivalent/Poised Promoter	H3K4me3/H3K27me3	Bernstein et al., 2006
Transcribed Gene Body	H3K36me3	Barski et al., 2007
Enhancer (both active and poised)	H3K4me1	Heintzman et al., 2007
Poised Developmental Enhancer	H3K4me1/H3K27me3	Creyghton et al., 2010; Rada-Iglesias et al., 2011
Active Enhancer	H3K4me1/H3K27ac	Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011
Polycomb Repressed Regions	H3K27me3	Bernstein et al., 2006; Lee et al., 2006
Heterochromatin	H3K9me3	Mikkelsen et al., 2007

survey DNA enrichment at ~500 representative loci using the nCounter probe system. High-quality antibodies are distinct from IgG patterns and the occupancy distribution correlates with a logical set of chromatin states. Validating antibody reagents for enrichment specificity and robustness ensures good quality ChIP-seq data sets with high signal to noise ratios.

Given that ChIP-seq is a mature technology, the technical restrictions of the technique are well defined by its users. These restrictions include the need for large amounts of starting material, limited resolution, and the dependence on antibodies. Improvements to ChIP-seq have been developed to address these limitations and expand the possibilities of its use. Collecting enough starting material for ChIP-seq can be challenging because experiments typically require 1 million (histone modifications) to 5 million (TFs and chromatin modifiers) cells. Although this is feasible when studying fast dividing cell lines, the challenge arises when studying primary cells and rare populations such as cancer stem cells or progenitor cells. ChIP-seq samples of 50,000 cells or less are possible with the ChIP-nano protocol (Adli and Bernstein, 2011). Key method modifications achieve effective chromatin fragmentation in small volumes, ensure minimal sample handling and loss by washing samples in columns, and reduce background signal. Another procedure, called ChIP-exo, improves the limited resolution from fragmentation heterogeneity after chromatin is prepared by sonication (Rhee and Pugh, 2011). As its name suggests, sonicated and immunoprecipitated DNA is treated with a 5'-to-3' exonuclease to digest DNA to the footprint of the crosslinked protein such that sequencing results are nucleotide resolution. This type of high-resolution protein-binding data is most beneficial for uncovering motifs of specific binding proteins and the effect of sequence variants on protein-binding affinity. Profiling genome-wide DNA-protein interactions with ChIP-seq is technically challenging when studying novel proteins or protein isoforms, such as a histone variant, that lacks a robust or specific antibody. In this case, an obvious approach is to transiently or stably express a protein of interest (POI) with a tag or epitope that can be readily ChIP'ed. Controls are necessary to ensure the fusion protein's localization is not altered by nonendogenous expression levels, protein instability, steric interference, or other effects of the tag itself.

A ChIP step can be added to other genomic profiling approaches for integrated epigenomic profiling. First, two ChIP

steps in a row, or Sequential-ChIP-seq, can uncover histone PTMs on the same molecule or chromatin-associated proteins in the same complex. Several groups combined bisulfite sequencing with ChIP giving rise to BisChIP-seq and ChIP-BS-seq (Brinkman et al., 2012; Statham et al., 2012). Long-distance DNA interactions mediated by a specific protein can be profiled using chromatin interaction analysis by paired-end-tag sequencing, or ChIA-PET (Fullwood et al., 2009). We anticipate other inventive uses of ChIP technology to continue to uncover undiscovered roles of histone modifications and histone variants in transcriptional regulation.

## Mapping of Chromatin Structures

### Nucleosome Positioning

Moving up the hierarchy of genomic organization, we now look beyond the DNA and histone modifications to the positioning of nucleosomes along the genome. Our epigenome at its most basic level is repeating units of 147 base pairs wrapped 1.7 times around each nucleosome with varying distances of linker DNA between each unit. Even this extremely simplistic model is complex because nucleosome positioning can both inhibit and promote factor binding (Bell et al., 2011). First, nucleosomes can be positioned to obstruct or reveal specific DNA sequences. Second, because modifications on histone tails serve as binding platforms for transcriptional regulators, nucleosome positioning regulates factor recruitment. And finally, nucleosomes are suggested to inhibit transcription by slowing progression of RNA polymerase II as it transcribes through a gene body. From a medical perspective, it will be important to determine the possible role of aberrant nucleosome positioning as caused by disease-associated SNPs, insertions, deletions, and translocations.

Our understanding of the regulation of nucleosome positioning came from studies of smaller genomes, such as those in yeast and fly (Jiang and Pugh, 2009). Nucleosome positioning along DNA is influenced by favorable DNA sequence composition, the actions of ATP-dependent nucleosome remodelers, and strongly positioned nucleosomes (Mavrich et al., 2008; Narlikar et al., 2013; Yuan et al., 2005). Although we understand the main determinants of nucleosome positioning, the exact contribution of each is unclear and currently under debate.

The most common method for profiling genome-wide nucleosome positioning is micrococcal nuclease digestion of

chromatin followed by high-throughput sequencing (MNase-seq). When native, uncrosslinked chromatin is digested with MNase, the linker DNA is cleaved while DNA wrapped around histone octamers or bound by TFs is protected. After purifying the DNA, approximately 150 bp fragments corresponding to mononucleosomes are size selected on a gel and sequenced.

Given the relative size of the human genome as compared to model organisms and that most of it is nucleosomal, mapping nucleosome positioning in the human genome is no small feat. Even with only 10-fold genome coverage in CD4<sup>+</sup> T cells, it was clear that nucleosomes are depleted at active transcriptional start sites (TSSs) and enhancers with ordered positioning radiating outward (Schones et al., 2008). In an extraordinary sequencing effort, Gaffney et al. generated paired and single-end MNase-seq data in seven lymphoblastoid cell lines yielding 240X coverage of a single cell type and found that ~80% of the genome has nonrandom, albeit weakly positioned nucleosomes (Gaffney et al., 2012). MNase-seq-based studies have provided great insights into the global distribution and dynamics of nucleosomes in the human genome. However, there are limitations of these data sets to consider. First, MNase digestion at the ends of nucleosomes is inconsistent such that exact position of nucleosomes can only be estimated as the center of fragments of different lengths. The results are unfortunately not single-base resolution. Moreover, MNase has AT sequence preferences, and MNase-protected 150 bp fragments are inferred to be mononucleosomes but could in fact be created by other proteins as well (Brogaard et al., 2012).

To circumvent the weaknesses of digestion-based detection of nucleosome positioning, Widom and colleagues used chemical modification of engineered histones to cleave DNA wrapped around nucleosomes (Brogaard et al., 2012). In short, DNA is precisely cleaved in a reaction with hydroxyl radicals if it interacts with a mutated residue on H4 while wound around a nucleosome. Using this method, nucleosome maps in yeast show remarkable accuracy and consistency. Interestingly, single-base-pair resolution of yeast nucleosomes reveals a 10 bp periodic sequence preference of flexible dinucleotides throughout the 147 bases in contact with the histone octamer. These data suggests that the role of sequence composition in the rotation positioning of the nucleosomes is stronger than previously appreciated. This method could be extended to the human genome for precise, high-resolution nucleosome positioning studies.

Another MNase-independent method for mapping nucleosome positioning uses DNA methyltransferase accessibility to footprint nucleosome positions, called nucleosome occupancy and methylome sequencing (NOME-seq) (Kelly et al., 2012). The unique activity of the DNA methyltransferase M.CviPI, which methylates cytosine only in the GpC context, is exploited to record the DNA's nucleosomal status because nucleosomal nucleotides are protected from methylation. Next, the DNA is bisulfite treated to convert unmethylated cytosines to thymine. Following sequencing, cytosines in the CpG context were originally methylated (5mC or 5hmC) and cytosines in the GpC context were nucleosome depleted. An important advantage of NOME-seq is the dual epigenomic information, both nucleosome position and DNA methylation abundance, comes from a single molecule rather than possibly co-occurring in a population of cells.

### Chromatin Accessibility

Nucleosomes are the basic repeated structural unit of the genome and proteins that compete for DNA binding can affect their positioning. Biochemically active regulatory elements, including promoters, enhancers, silencers, and insulators, are bound by sequence-specific regulatory TFs. Open chromatin is therefore an overarching characteristic of biochemically active genomic regions. Open chromatin can be assayed genome-wide by DNase hypersensitivity followed by sequencing (DNase-seq) or formaldehyde-assisted identification of regulatory elements followed by sequencing (FAIRE-seq) (Boyle et al., 2008; Giressi et al., 2007). DNase-seq takes advantage of the protection conferred by tightly wound nucleosomes from DNaseI endonuclease digestion. Accordingly, limited digestion of native chromatin releases nucleosome-depleted fragments. Sequencing and mapping of these fragments identifies DNaseI-hypersensitive sites (DHSs) corresponding to regulatory regions. Protein and complex binding within a DHS create 6–40 bp sequences of DNaseI protection, called digital DNase footprints, which can be called after deep sequencing of DNase-seq libraries (Neph et al., 2012b). Beyond being DNaseI hypersensitive, open chromatin regions are also sensitive to shearing by sonication, and this concept is exploited by the FAIRE-seq assay. Initially, chromatin isolated from formaldehyde crosslinked cells is subject to sonication. Subsequently, free DNA unbound by proteins is isolated from the aqueous phase following phenol-chloroform extraction. Similar to DNase-seq, mapping the FAIRE-seq enriched fragments demarcate open chromatin and regulatory elements.

Myriad chromatin accessibility maps are now available due to the adoption of DNase-seq by the ENCODE Project Consortium and Roadmap Epigenomics Program. In total, these consortia produced hundreds of maps encompassing 349 cell types including pluripotent stem cells, stem cell progenitors, cultured primary cells, and human fetal tissue from various types and gestational stages (Maurano et al., 2012). Meta-analysis of these data sets determined DHSs span 2.1% of the genome per cell type on average, and, impressively, all ~4,000,000 sites collectively cover ~40% of the genome (Maurano et al., 2012). Mapping nucleosome-depleted chromatin is a comprehensive way to identify the global catalog of regulatory elements and factor-binding sites without specific antibodies or prior knowledge of cell-type-specific transcriptional regulators.

### Mapping of Higher Order Chromatin Architecture

Lower-order chromatin structures such as the 11 nm fiber, also called “beads on a string,” are followed by higher-order structures like the 30 nm fiber and 700 nm mitotic chromosomes (Felsenfeld and Groudine, 2003). Incidentally, genome compaction brings regions that are linearly distant via short and long-range chromatin interactions. Mapping the 3D structure of the nucleus is important because, like histone marks and chromatin accessibility, chromosome conformations influence mammalian gene regulation. For example, chromatin in close proximity to the lamina of the inner nuclear membrane, or lamin-associated domains (LADs), tends to be heterochromatic and transcriptionally repressed (Akhtar et al., 2013; Guelen et al., 2008). In contrast, high local concentrations of RNA polymerase II bound

promoters, called transcription factories, correlate with robust gene expression (Brown et al., 2008). Historically, fluorescent microscopy imaging, though limited by resolution and throughput, has served as the gold standard for observing nuclear structure. In fact, both LADs and transcription factories were discovered in this fashion.

A paradigm shift from FISH-based methods toward measuring physical DNA interactions has revolutionized our ability to broadly map nuclear architecture. The introduction of chromosome conformation capture (3C) by Dekker and colleagues provided an alternative to mapping distances between two loci (Dekker et al., 2002). Succinctly, chromatin from formaldehyde crosslinked cells is digested with a restriction enzyme followed by DNA ligation under extremely dilute conditions to favor joining of ends in close proximity to each other. After reversal of crosslinks and DNA purification, the ligation frequency between two restriction fragments, measured by qPCR, indicates their interaction frequency. Such interaction frequency is generally related to the spatial distance, though this relationship could be complex, nonlinear, and influenced by chromatin accessibility (Dekker et al., 2013; Williamson et al., 2012).

Differing in the number of loci tested and selection of loci, the suite of 3C-based assays include 3C, circular chromosome conformation capture (4C), chromosome conformation capture carbon copy (5C), tethered conformation capture (TCC), ChIA-PET, and Hi-C (reviewed by de Wit and de Laat, 2012; Sajan and Hawkins, 2012; van Steensel and Dekker, 2010). Of these, 3C and 5C are locus-centric methods, meaning the assayed regions are selected a priori. Sequence-specific primers are designed for each locus of interest, for instance promoters or *cis*-regulatory elements. As stated previously, 3C measures the interaction frequency of one locus with another (one-to-one), between an anchor and a bait sequence, by qPCR. 5C attains higher throughput capacity by utilizing thousands of anchor and bait primers that can span an entire chromosome (many-to-many) (Dostie et al., 2006). The 4C method measures the genome-wide interaction frequency of a single anchor site. Inverse PCR of ligated and circularized interacting DNA fragments detects all interacting loci (one-to-all) (Zhao et al., 2006).

Hi-C measures the entire genome's interaction frequency with itself as a large matrix without enrichment for specific loci (all-to-all) (Lieberman-Aiden et al., 2009). The key innovation of Hi-C is the ability to enrich for ligation junctions of interacting fragments. After digestion of crosslinked chromatin with a 6-base cutter, the overhang ends are filled in with a biotinylated base. As with all 3C-based assays, the DNA fragments are extremely dilute in the presence of ligase to promote intramolecular ligation events. The ligated sample is sonicated and a streptavidin pulldown captures all junctions of interacting DNA fragments, which are then sequenced. Hi-C interaction signal-to-noise ratio is dependent on the rate of intramolecular ligation events. TCC, a Hi-C variation, performs the ligation step on the streptavidin bead to promote these favorable ligation events (Kalhor et al., 2012). ChIA-PET is another variation of Hi-C that features an immunoprecipitation step to map DNA interactions involving a POI (Fullwood et al., 2009). This approach has been useful for understanding proteins involved in nuclear organiza-

tion such as TFs, RNA polymerase, CTCF, and cohesin (Demare et al., 2013; Fullwood et al., 2009; Handoko et al., 2011; Li et al., 2012).

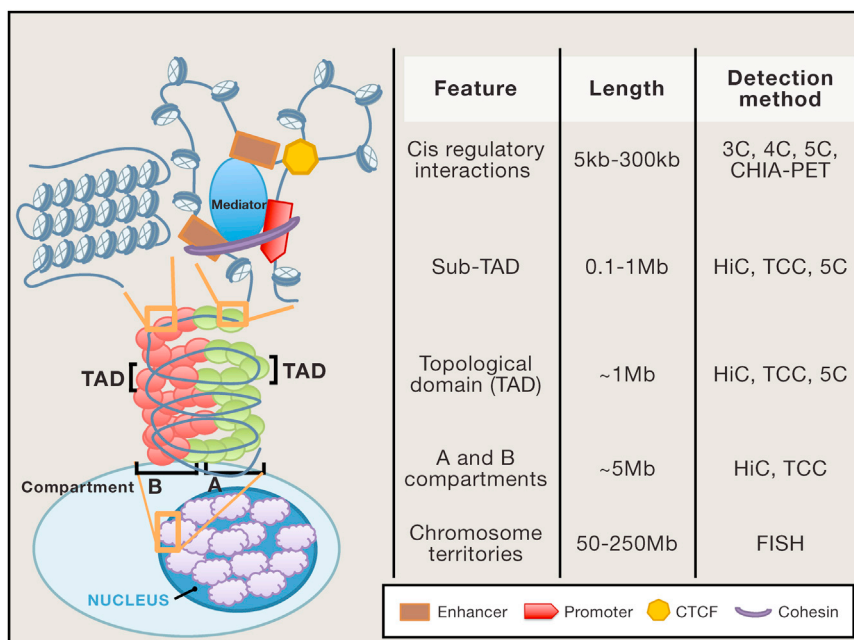
Due to its unbiased genome-wide scale, Hi-C promotes the discovery of novel interactions and structures. The size or level of nuclear architecture assayed with Hi-C is dependent on experimental resolution, which in turn is limited by both restriction enzyme cutting frequency and sequencing depth. At 1 Mb resolution, Dekker and colleagues defined higher-order chromosome compartments that average about 5 Mb in size across the genome (Lieberman-Aiden et al., 2009). Termed A and B compartments, these sequentially interchanging structures correlate with euchromatic and heterochromatic genomic regions, respectively, in a cell-type-specific manner. With increased sequencing depth, we attained 40 kb resolution Hi-C interaction data. In this way, our lab identified cell-type and species invariant 1 Mb regions of high local interaction frequency, called topological domains or topologically associated domains (TADs), separated by noninteracting boundary elements (Dixon et al., 2012). Boundary regions correlate with the limits of heterochromatin blocks and are enriched for CTCF-binding sites, housekeeping genes, and certain transposon elements. With deeper sequencing and improved data analysis algorithms for Hi-C data, it may be possible to reach fragment length resolution or about 4 kb when using a 6-base cutter.

What has become clear using microscopy and now 3C-based assays is a gradient or spectrum of DNA folding architecture (Figure 3). These interactions start as one-to-one looping of regulatory elements including enhancers, promoters, and insulators. A collection of these interactions flanked by boundary regions form local interacting neighborhoods or TADs (Dixon et al., 2012; Nora et al., 2012). Using 5C, sub-TADs were recently recognized as finer resolution segments of TADs that can be cell-type-specific (Phillips-Cremins et al., 2013). Together many TADs contribute to a larger chromosome compartment and, finally, to an entire chromosome territory (Lieberman-Aiden et al., 2009). In summary, the advent of 3C technologies has revealed novel and complex layers of genome annotation. Many questions remain regarding how nuclear architecture is regulated and how it mechanistically influences transcription, cellular identity, and possibly disease.

### Utility of Epigenome Maps

In addition to rapid development of new epigenome mapping technologies, another important factor fueling the remarkable progress of epigenomics is the formation of many research consortia worldwide. Most notable among these consortia are the US Roadmap Epigenome Project, the ENCODE Project, and the International Human Epigenome Consortium (IHEC) (Table 1). Modeling after the hugely successful Human Genome Project (Lander, 2011), these consortia standardized experimental protocols, recommended data analysis procedures, and most importantly, publicly released large amounts of data sets prior to publication. Consequently the number of epigenome maps generated has grown exponentially, from a handful in early 2007 to several thousand as of today. As detailed below, these maps have proved to be most valuable in annotation of the transcription units, *cis*-regulatory sequences, and other genomic





**Figure 3. Hierarchical Principles of Nuclear Organization**

man et al., 2009). Similarly, our lab showed that transcriptional enhancers are characterized by the presence of H3K4me1 but not H3K4me3, a combination that accurately predicted tens of thousands of new enhancers in the human genome (Heintzman and Ren, 2009; Heintzman et al., 2007). Despite initial hesitation that this signature may not universally demarcate enhancers in different cell types or in more than one species, H3K4me1 marks enhancers in all tested cell types from human to zebrafish to fly (Aday et al., 2011; Heintzman and Ren, 2009; Roy, et al., 2010). Importantly, as researchers have annotated enhancers in many human cell types, they have realized that cell-type specificity is a persistent feature of enhancers

and is consistent with their role in determining cellular identity. Other general properties of enhancers that are validated genome-wide include DNaseI hypersensitivity, combinatorial TF binding, H3.3 and H2A.Z histone variant enrichment, bound RNA Pol II, and RNA production (enhancer RNAs or eRNAs) (Buecker and Wysocka, 2012). Combinations of these features along with H3K4me1 allow refined and highly accurate genome-wide enhancer prediction (Rajagopal et al., 2013).

### Annotation of *cis*-Regulatory Elements from Chromatin Profiles

A major challenge confronting biomedical researchers is the absence of functional annotation of the human genome, where the vast majority (98.5%) do not code for proteins yet harbor most of the disease-associated genetic variations. Several types of functional sequences are known to exist in the noncoding parts of the human genome including *cis*-regulatory elements. These sequences, including promoters, enhancers, and insulators, govern gene expression by recruiting sequence-specific TFs that modulate local chromatin structure and assembly of transcriptional machinery. Because these functional DNAs lack consistent and recognizable sequence features, their identification and characterization had until recently been quite difficult.

Epigenome maps have proven to be a powerful tool for annotating the functional features in the human genome. This is possible because several classes of DNA elements are associated with characteristic histone modifications (Table 3). Early studies revealed that H3K4me3 predominantly associates with promoters and H3K36me3 with gene bodies. Thus, new gene units could be identified with the use of chromatin profiles. Indeed, this strategy led to the identification of several thousands of long noncoding intergenic RNA genes (lincRNA) (Gutt-

man et al., 2009). Similarly, our lab showed that transcriptional enhancers are characterized by the presence of H3K4me1 but not H3K4me3, a combination that accurately predicted tens of thousands of new enhancers in the human genome (Heintzman and Ren, 2009; Heintzman et al., 2007). Despite initial hesitation that this signature may not universally demarcate enhancers in different cell types or in more than one species, H3K4me1 marks enhancers in all tested cell types from human to zebrafish to fly (Aday et al., 2011; Heintzman and Ren, 2009; Roy, et al., 2010). Importantly, as researchers have annotated enhancers in many human cell types, they have realized that cell-type specificity is a persistent feature of enhancers

and is consistent with their role in determining cellular identity. Other general properties of enhancers that are validated genome-wide include DNaseI hypersensitivity, combinatorial TF binding, H3.3 and H2A.Z histone variant enrichment, bound RNA Pol II, and RNA production (enhancer RNAs or eRNAs) (Buecker and Wysocka, 2012). Combinations of these features along with H3K4me1 allow refined and highly accurate genome-wide enhancer prediction (Rajagopal et al., 2013).

Initially, chromatin states in two human cell lines, HeLa cells and K562 cells, were mapped and used to predict 55,000 candidate enhancers in the human genome (Heintzman et al., 2009). This study demonstrated that chromatin modifications at enhancers, in particular H3K4me1 and H3K27ac, are cell-type-specific and correlate with cell-type-specific gene expression throughout the genome, suggesting a potentially critical role for enhancers in lineage-specific gene regulation. Subsequent studies confirmed this result in additional cell types. Rada-Iglesia and colleagues characterized the chromatin modification profiles in the human embryonic stem cells (hESC) and identified approximately 7,000 candidate enhancers featuring binding of TFs, presence of H3K4me1 mark, and depletion of nucleosomes (Rada-Iglesia et al., 2011). Interestingly, they also classified these enhancers into “active enhancers” and “poised enhancers,” which differ mainly in the presence or absence of H3K27ac mark. Active enhancers are near genes expressed in hESCs, while poised enhancers are next to genes inactive in hESC but turned on during differentiation (Rada-Iglesia et al., 2011). Independently, Creyghton and colleagues, by examining the chromatin modification patterns in the mESCs and several differentiated mouse cell types, also found that H3K27ac can distinguish the active enhancers from poised ones (Creyghton et al., 2010). While tens of thousands of enhancers are typically marked in a cell type, a small subset of enhancers (<1%) form large domains up to 50 kb marked by H3K4me1, H3K27ac,

mediator, and master transcription factor binding (Whyte et al., 2013). These “superenhancers” are posited to regulate key genes important for cell identity. For example, in mESCs superenhancers are associated with genes necessary for pluripotency and in myotubes they are associated with skeletal muscle development. Lastly, enhancers are turned off or “decommissioned” by the removal of H3K4me1 and H3K27ac marks when they no longer pertain to the cellular state. For example, enhancers that control genes important for pluripotency must be decommissioned for proper differentiation (Whyte et al., 2012). By profiling key chromatin marks in different cell types, researchers can now annotate the location and activity state of regulatory elements across the genome to define a cell-type’s regulome.

Extending the idea that chromatin modification patterns are associated with different functional sequences, a comprehensive epigenomic study of nine human cell lines showed that the human genome could be segmented into regions carrying one of 15 different combinations of chromatin modification marks in each cell type (Ernst et al., 2011). The authors implemented a multivariate hidden Markov model called ChromHMM to unbiasedly infer these chromatin states from the chromatin modification profiles. This approach found that each chromatin state corresponds to a specific category of genomic features, including active or poised promoters, enhancers, insulators, and silenced domains (Table 3). Similar results have been obtained with other machine-learning algorithms (Hoffman et al., 2013; Won et al., 2013). More recently, the ENCODE Consortium profiled the chromatin modification state in 46 human cell types, and found that as much as 56% of the genome is associated with specific histone modification patterns indicative of biochemical activities (ENCODE Project Consortium, et al., 2012). Ongoing studies from large epigenome mapping consortia such as the NIH Epigenome Roadmap consortium are certain to enhance human genome annotation even further (Bernstein et al., 2010).

### Annotation of Long-Range Chromatin Interactions

Rather than the regulome simply being a collection of elements, it must ultimately link together enhancers, insulators, promoters, and other features in three-dimensional space. Epigenomic maps can annotate cell-type-specific long-distance interactions between regulatory elements using two approaches. First, long-range looping can be mapped by physical interaction of distant DNA loci with 3C-based assays. Considering their capabilities, 5C and ChIA-PET currently provide the best balance of resolution and reasonable coverage in the human genome for this purpose (Dekker et al., 2013; Smallwood and Ren, 2013). Several 5C and ChIA-PET data sets exist that are valuable for identifying interactions in selected human cell types, but diverse interaction maps are not yet available. The second method correlates regulatory elements and target promoters across many cell types to computationally infer pairs that regulate and likely interact with each other. These *in silico* predictions require the annotation of cell-type-specific enhancers and active promoters with chromatin signatures, DHSs, and expression profiles (Ernst et al., 2011; Sheffield et al., 2013; Shen et al., 2012; Thurman et al., 2012). A clear advantage of the correlative approach is the necessary data in hundreds of cell types had already been produced by epigenomic consortia. For example, Thurman et al.

defined almost 600,000 regulatory pairs from 79 cell types by correlating ENCODE DNase-seq and RNA-seq data (Thurman et al., 2012). Some of the predicted interaction correlations defined by this approach were validated by both 3C and ChIA-PET but more extensive cross-validation is necessary. Furthermore, such results are limited to enhancer-promoter communication and cannot address the role of insulators or novel interacting regions.

Although these two approaches use different concepts to define chromatin interactions, they independently come to similar conclusions about nuclear architecture. The main theme is consistent with the role of enhancers in the determination of cellular identity: enhancer-promoter interactions are cell-type-specific (Sanyal et al., 2012; Shen et al., 2012). As previously suggested from single-locus studies, enhancers target promoters at long distances and do not always target the nearest gene (Ong and Corces, 2011). This is confirmed by a broader study using 5C analysis of 1% of the human genome in three cell types that determined that ~50% of distal regulatory elements interact with the closest active gene (Sanyal et al., 2012). Surprisingly, enhancers and promoters can interact promiscuously with more than one element indicating a complicated web of transcriptional regulation (Sanyal et al., 2012; Shen et al., 2012; Thurman et al., 2012). In fact, Thurman et al. used a correlative approach and found half of TSSs regulate ten or more distal sites and half of putative enhancers regulate more than one TSS (Thurman et al., 2012). 5C physical interaction data also confirms 49% of TSSs interact with more than one distal site but finds only 10% of enhancers interacting with more than one promoter (Sanyal et al., 2012). Another concept challenged by interactome annotations, is the classic definition of CTCF-bound insulators as having an enhancer-blocking function to limit the range of targeted enhancer activation (Phillips and Corces, 2009). Notably, enhancer-promoter interactions often span hundreds of kilobases surpassing one or more CTCF sites (Demare et al., 2013; Sanyal et al., 2012). 5C analysis predicts that almost 60% of all interactions skip a cobound CTCF/cohesin locus (Sanyal et al., 2012). Another unexpected result is frequent insulator interactions with promoters and enhancers suggesting a possible role for CTCF in transcriptional activation (Demare et al., 2013; Handoko et al., 2011; Sanyal et al., 2012). Finally, there is an emerging role of CTCF and cohesin cobound regulatory elements in long-range, constitutive DNA interactions in contrast to cohesin and mediator cobound regions that facilitate short range, cell-type-specific enhancer-promoter interactions (Demare et al., 2013; Phillips-Cremens et al., 2013; Sanyal et al., 2012).

On the whole, great progress has been made toward understanding how chromatin and nuclear organization jointly regulate global gene expression patterns. Still, it remains to be determined how up to 20 enhancers choose to target a particular promoter, what the functional consequence is for a promoter to be activated by more than once enhancer, and if these results are merely an artifact of averaging interactions and epigenetic maps over a population of cells. Evidence that enhancer interactions frequently bypass insulators and that CTCF can loop to transcriptional start sites suggests we must refine our understanding of CTCF-bound insulators. Certainly the conclusions

from these epigenomic studies will encourage mechanistic studies to uncover the details of transcriptional regulation by long-range interactions. Continuation of this work at high resolution and in many cell types will progress the dimensions of chromatin maps from a linear chart to a three-dimensional model of genomic annotation.

### Dynamic Chromatin Landscapes during Human Development

Advances in the robustness, throughput, and accuracy of epigenomic technologies over the past decade have co-occurred with a revolution in stem cell biology and regenerative medicine. The first completed epigenomes documented the features of cultured, immortalized or cancerous cell lines. Although this enabled correlation of epigenomic features with genomic elements, the resulting epigenomic landscapes were a static view. Now that lineage-specific differentiation protocols produce populations of cells with increasing purity, it is possible to study epigenomic dynamics during development. Such studies give insight into how cell state transitions are influenced by chromatin states at promoters and enhancers, DNA methylation dynamics at enhancers, and the expansion of repressive domains. Finally, using chromatin signatures to define lineage-specific enhancers can lead to models of TF networks that regulate cell-type specification.

Unique chromatin signatures mark promoters and enhancers in pluripotent cells. Some promoters in ESCs are bivalent or comarked with active and repressive histone marks. Sequential ChIP confirmed the presence of both H3K4me3 and H3K27me3 on the same DNA molecule or promoter allele (Bernstein et al., 2006). Due to polycomb-mediated silencing, bivalent genes are not expressed and specifically mark developmental genes that are activated in downstream cell states and are “poised” in ESCs. Consistent with this idea, differentiation of hESCs into endoderm, mesoderm, and ectoderm lineages resolves 85% of bivalent promoters into monovalent states in a lineage-specific manner (Gifford et al., 2013). H3K4me1/H3K27me3 marked poised enhancers are a special class found mostly in ESCs and tend to regulate bivalent promoters of developmental genes (Rada-Iglesias et al., 2011). Chromatin state maps can have more informative power than expression data alone, because they identify both active genes and poised genomic elements that foreshadow a cell type's differentiation potential.

DNA methylation can also be informative for identifying enhancers and classifying their activity. Quantitative comparisons of the first human methylomes in hESCs and fibroblasts indicated regions with dynamic DNA methylation levels. A region that is relatively hypermethylated in one cell type and hypomethylated in another is called a differentially methylated region (DMR) (Lister et al., 2009). DMRs are enriched at regulatory elements as evidenced by their overlap with DNaseI sites, TF-binding sites, and enhancer chromatin marks (Hon et al., 2013; Ji et al., 2010; Ziller et al., 2013). Active enhancers correlate with the hypomethylated DMR state, also called low-methylated regions (LMRs), and the formation and maintenance of LMRs is dependent on TF binding (Stadler et al., 2011). However, whether the depletion of methylation upon TF binding occurs passively during cell cycles or through active demethylation is

still unclear. Recently generated methylomes of cytosine variants find 5hmC, 5fC, and 5caC enriched at enhancers in ESCs suggesting the latter is possible (Pastor et al., 2011; Shen et al., 2013; Szulwach et al., 2011; Yu et al., 2012). Even inactive enhancers can be identified by hypomethylation. We recently discovered “vestigial enhancers,” defined as regions depleted of DNA methylation and enhancer chromatin marks in adult tissues but which exhibit enhancer activity earlier in development (Hon et al., 2013). These results suggest that the methylome retains a memory of its previous cellular identities, although how and why is still unclear. Overall, dynamic DNA methylation patterns at enhancers provide yet another epigenetic feature for their annotation. In fact, unexpected regions of hypomethylation in transposable elements helped to annotate these regions as tissues-specific enhancers (Xie et al., 2013a).

Cellular differentiation during development is characterized by gradual expansion of repressed domains. In ESCs, H3K27me3 is distributed across 8% of the genome and is most commonly observed at bivalent promoters of developmental genes (Zhu et al., 2013). First observed in IMR90 fibroblasts but since confirmed in many differentiated cell types, H3K27me3 peaks spread from bivalent promoters in ESCs to form large blocks covering up to 40% of the genome in differentiated cells (Hawkins et al., 2010; Zhu et al., 2013). These large repressed regions specifically span developmental genes, and presumably function to silence genes that do not pertain to the specified lineage. Correspondingly, chromatin accessibility at regulatory DNA is progressively lost during differentiation (Stergachis et al., 2013). The epigenome of ESCs is uniquely open and accessible, which is consistent with the role of these cells as pluripotent progenitors of all three germ layers. It is now clear that during differentiation and development, the epigenome is progressively restricted like its plasticity potential.

Another major utility of epigenomic studies is the ability to identify cell-state- or stage-specific master regulators and construct transcriptional networks. DNase-seq-based methods are useful for mapping all TF-binding sites genome-wide in one assay. DNase footprint analysis in 41 human cell types from ESCs to primary adult cells identified 45 million footprints and subsequent de novo motif analysis discerned over 600 unique motifs (Neph et al., 2012b). This set of motifs recovered 90% of the TRANSFAC, JASPAR, and UniPROBE motif database entries. However, nearly half of the motifs found in this study are novel and bound by an unknown protein indicating there is a substantial amount of work to be done to have a complete understanding of sequence-dependent transcriptional regulators. Similar to this approach, master transcriptional regulators can be identified by annotating putative enhancers using histone modification signatures followed by motif analysis of these regions. In this way, novel master regulators of human neural crest (NR2F1), cardiac development (Meis2), and adipogenesis (PLZF) have been successfully identified (Mikkelsen et al., 2010; Paige et al., 2012; Rada-Iglesias et al., 2012). Enhancer-driven regulatory networks can be constructed by correlating TF expression data with motif analysis at putative enhancers as predicted by histone marks or open chromatin (Ernst et al., 2011; Neph et al., 2012a). Alternatively, TF networks can be directly assayed using ChIP-seq. The ENCODE consortium

characterized the binding patterns of 119 TFs in five cell lines by ChIP-seq that yielded ordered predictions of TF hierarchies, rules of combinatorial TF binding, and TF-binding preferences for distal or proximal sites (Gerstein et al., 2012). However with over 1,000 TFs and many cell types still not assayed, the complexity of experimentally and computationally generated transcription factor regulatory networks will increase and requires further investigation.

### Understanding Disease Variants with Epigenomics

The combination of data from the Human Genome Project and International HapMap Project enabled geneticists to study polygenic traits and diseases using GWAS (Frazer et al., 2009). Investigating the genetic basis of common diseases is possible by testing for variants that significantly associate with cases over controls in a population. GWAS are now a common tool in genetic epidemiology with over 1,600 studies published associating 11,000+ SNPs with hundreds of distinct diseases and traits (Hindorf and MacArthur, 2013). It is often difficult when interpreting GWAS results to identify the functional variant(s) underlying an association. The difficulty arises from the fact that a SNP associated with a particular phenotype may itself be the functional variant directly causing or contributing to the phenotype, or it may simply be in linkage disequilibrium (LD) with the functional variant. Traditional GWAS alone cannot differentiate between these two scenarios. Moreover, even after identifying a true functional variant, it can be very difficult to determine the molecular mechanism by which that variant alters genome function and contributes to disease. For example, 93% of SNPs associated with human phenotypes by GWAS are located outside of protein coding regions (Maurano et al., 2012). This finding highlights the need for annotation of the vast noncoding sequence of the human genome. Epigenomic maps are likely to play a major role in identifying specific functional variants that cause or contribute to human disease.

Assimilation of GWAS and epigenomic data can give clues about the involvement of regulatory elements and their target genes in the pathogenesis of common disease. Annotating enhancers by their chromatin state in nine cell types, Ernst et al. found that disease-associated SNPs are enriched within cell-type-specific enhancers from the appropriate disease cell type (Ernst et al., 2011). By correlating the locations of over 5,000 non-coding common disease-associated SNPs with almost 4 million DHSs from hundreds of cell types, Maurano et al. found 40% of GWAS SNPs fall within these regions (Maurano et al., 2012). Due to the cell-type specific property of regulatory elements, correlation of all GWAS SNPs in DHSs from a specific disease can facilitate de novo prediction of pathogenic tissues. Using this approach, GWAS SNPs associated with Crohn's disease are enriched in the cell-type specific DHSs of TH17 and TH1 cells, and both of these cell types have been independently associated with this disease (Maurano et al., 2012).

Variants can affect both local open chromatin and gene expression and these loci are called DNase sensitivity quantitative trait loci (dsQTLs) (Degner et al., 2012). Identification of dsQTLs in haplotyped lymphoblastoid cells from 70 individuals found dsQTLs frequently fall within TF-binding sites, affect nucleosome positioning as determined by MNase-seq, and

overlap with GWAS disease-associated SNPs (Degner et al., 2012; Gaffney et al., 2012). Interestingly 80% of dsQTLs lie in chromatin regions previously predicted to be functional in lymphoblastoid cell lines by epigenetic marks, including 41% in predicted enhancers, suggesting they may be the functional variant.

Reference epigenomes in diverse tissues and the annotation of millions of putative enhancers can help researchers make hypotheses about which SNPs in LD with GWAS SNPs may contribute to disease. Following up GWAS studies by genotyping regulatory elements around GWAS SNPs may more definitively identify functional SNPs. Importantly, disease-associated SNPs that reside in putative regulatory elements must be functionally validated. SNPs in enhancers may affect their activity by disrupting TF binding, coactivator recruitment, looping interactions with target promoters, or transcription of the target promoter. To this end, each of these possible outcomes can be tested with traditional approaches such as reporter assays, gel shift assays, and allele-specific ChIP. In addition, emerging technologies such as TALENs, CRISPR/Cas9, 3C-based technologies, and new massively parallel reporter assays promise to vastly improve the rigor and throughput with which we can test these hypotheses (Melnikov et al., 2012; Patwardhan et al., 2012).

### The Road Ahead

Six decades ago, Watson and Crick put forward a model of DNA double helix structure to elucidate how genetic information is faithfully copied and propagated during cell division (Watson and Crick, 1953). Several years later, Crick famously proposed the "central dogma" to describe how information in the DNA sequence is relayed to other biomolecules such as RNA and proteins to sustain a cell's biological activities (Crick, 1970). Now, with the human genome completely mapped, we face the daunting task to decipher the information contained in this genetic blueprint. Twelve years ago, when the human genome was first sequenced, only 1.5% of the genome could be annotated as protein coding, whereas the rest of the genome was thought to be mostly "junk" (Lander et al., 2001; Venter et al., 2001). Now, with the help of many epigenome maps, nearly half of the genome is predicted to carry specific biochemical activities and potential regulatory functions (ENCODE Project Consortium, et al., 2012). It is conceivable that in the near future the human genome will be completely annotated, with the catalog of transcription units and their transcriptional regulatory sequences fully mapped.

However, to reach this goal a couple of technical issues should be resolved. First, current approaches to epigenomic analysis still demand a large number of cells, limiting the cell types and developmental stages that can be examined. Robust nanoscale techniques for chromatin modification profiling, methylcytosine detection or chromatin accessibility mapping will be necessary to cover the full spectrum of cellular states. A few nanoscale ChIP-seq, whole-genome bisulfite sequencing methods have been reported, enabling epigenomic analysis of rare cell populations such as the germ cells (Adey and Shendure, 2012; Adli and Bernstein, 2011; Ng et al., 2013). We expect that the application of these methods to the rare cell types and to embryonic stages



will significantly broaden our knowledge of the dynamic human epigenome, and facilitate the discovery of additional cell-type-specific regulatory elements and transcription units. Second, a lot of the epigenome maps currently available are from tissues consisting of heterogeneous cell populations, and do not accurately reflect the epigenomic state of a specific cell type. Today, to isolate individual cell types from a tissue in large quantity for epigenomic analysis by FACS is challenging. With the advance in nanoscale epigenomic analysis technique, it will likely be routine to obtain cell types in small quantity without affecting their cellular state and produce epigenome maps for specific cell populations in a tissue type. Eventually, the solution will likely be single-cell epigenome analysis techniques. Currently, it is not yet possible to profile the chromatin state or DNA methylation status of a single cell, but this could become a reality with further development of single-molecule techniques such as SCAN (single chromatin molecule analysis in nanochannels) (Murphy et al., 2013) or SMRT (single-molecule, real-time) sequencing (Clarke et al., 2009; Flusberg et al., 2010).

Ultimately, to truly understand how the genome programs human development and how certain sequence variants cause human disease, we ought to be able to predict from the sequence when and at what level a gene is expressed in different cell types. To achieve this goal, we need to overcome substantial challenges in two areas. First, we must gain a better understanding of the functional relationships between DNA methylation, chromatin modification state, or higher order chromatin structure and gene regulation. Epigenomic studies have established correlations between DNA hypomethylation, certain chromatin modification signature, and chromatin structure at *cis*-regulatory elements such as enhancers and promoters, but to determine whether an epigenetic state is necessary for transcription or merely coincidental will require additional mechanistic studies. To this end, new technologies are required to allow manipulation of the epigenetic state of specific loci. The recent development of TALE factors and CRISPR/Cas9 as a way to target histone modification enzymes or DNA methyltransferases to specific genomic sites is a great example (Gaj et al., 2013; Ramalingam et al., 2013; Wang et al., 2013). Second, we need to better define the target genes of the distal regulatory elements such as enhancers. This is a challenge because many enhancers are located hundreds and thousands of base-pairs away from their target genes, and it is not infrequent that an enhancer and its target gene are separated by other irrelevant genes (Smallwood and Ren, 2013). Techniques such as ChIA-PET, 4C, 5C, and Hi-C have been invented to identify looping interactions between enhancers and target promoters (de Wit et al., 2013; Li et al., 2012; Sanyal et al., 2012; Wei et al., 2013). However, it is currently unclear whether mere spatial proximity is sufficient for functional regulation. Thus, mapping physical interactions alone is unlikely to be adequate to resolve the target genes for enhancers. An alternative, complementary approach defines target genes for an enhancer by searching for nearby genes sharing similar chromatin state or accessibility across diverse cell types (Ernst et al., 2011; Shen et al., 2012; Thurman et al., 2012). Despite its simplicity, this approach does not work particularly well when a gene is regulated by multiple tissue-specific enhancers in different tissues. A hybrid approach combining both spatial

proximity and chromatin state information will likely more accurately define the target genes for enhancers.

In summary, the last few years have witnessed an explosion of the epigenomic field. Thousands of epigenome maps from hundreds of human cell or tissue types have been produced, adding crucial insights into the second dimension of the genomic sequence. Although these human epigenomes have illuminated potentially functional elements in the genome and improved our understanding of the human developmental programs, it is also clear that much more remains to be explored. Better characterization of epigenome variations in human populations and in patients will be critical for us to fully appreciate the epigenetic factors in human health and disease etiology. Indeed, projects are already underway to profile thousands more epigenomes from both healthy and diseased individuals (American Association for Cancer Research Human Epigenome Task Force; European Union, Network of Excellence, Scientific Advisory Board, 2008; Bernstein et al., 2010). We anticipate an even greater revolution in our understanding of the human epigenome in the coming years.

#### ACKNOWLEDGMENTS

We apologize to those authors whose works are not covered here owing to space limitations. We thank Peter Jones, Bradley Bernstein, Fred Tyson, John Satterlee, and members of the Ren Lab for their comments on earlier versions of this manuscript. This work is supported by funds from the Ludwig Institute for Cancer Research, NIH (grants U01ES017166 and U54 HG006997), and the California Institute for Regenerative Medicine (RN2-00905) to B.R. C.M.R. was supported in part by the UCSD Genetics Training Program through an institutional training grant from the National Institute of General Medical Sciences, T32 GM008666.

#### REFERENCES

- Aday, A.W., Zhu, L.J., Lakshmanan, A., Wang, J., and Lawson, N.D. (2011). Identification of *cis* regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. *Dev. Biol.* 357, 450–462.
- Adey, A., and Shendure, J. (2012). Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. *Genome Res.* 22, 1139–1143.
- Adli, M., and Bernstein, B.E. (2011). Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nat. Protoc.* 6, 1656–1668.
- Akhtar, W., de Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F.A., van Lohuizen, M., and van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* 154, 914–927.
- American Association for Cancer Research Human Epigenome Task Force; European Union, Network of Excellence, Scientific Advisory Board. (2008). Moving AHEAD with an international human epigenome project. *Nature* 454, 711–715.
- Badeaux, A.I., and Shi, Y. (2013). Emerging roles for chromatin as a signal integration and storage platform. *Nat. Rev. Mol. Cell Biol.* 14, 211–224.
- 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.
- Beck, S., Bernstein, B.E., Campbell, R.M., Costello, J.F., Dhanak, D., Ecker, J.R., Gready, J.M., Issa, J.P., Laird, P.W., Polyak, K., et al.; AACR Cancer Epigenome Task Force. (2012). A blueprint for an international cancer epigenome consortium. A report from the AACR Cancer Epigenome Task Force. *Cancer Res.* 72, 6319–6324.

- Bell, O., Tiwari, V.K., Thomä, N.H., and Schübeler, D. (2011). Determinants and dynamics of genome accessibility. *Nat. Rev. Genet.* 12, 554–564.
- Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169–181.
- 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, 315–326.
- Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. *Cell* 128, 669–681.
- Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* 28, 1045–1048.
- Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. *Genomics* 98, 288–295.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.
- Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.* 13, 705–719.
- Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. *Science* 330, 612–616.
- Booth, M.J., Branco, M.R., Ficiz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336, 934–937.
- 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.
- Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., and Stunnenberg, H.G. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* 22, 1128–1138.
- Brogaard, K., Xi, L., Wang, J.-P., and Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. *Nature* 486, 496–501.
- Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A., Smith, A.J., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R., et al. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* 182, 1083–1097.
- Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* 28, 276–284.
- Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol. Cell* 49, 825–837.
- Clark, S.J., Harrison, J., Paul, C.L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22, 2990–2997.
- Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* 4, 265–270.
- Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215–219.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* 107, 21931–21936.
- Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563.
- de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24.
- de Wit, E., Bouwman, B.A.M., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J.A.M., Krijger, P.H.L., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*. Published online July 24, 2012.
- Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.-B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNase-seq sensitivity QTLs are a major determinant of human expression variation. *Nature* 482, 390–394.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311.
- Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* 14, 390–403.
- Demare, L.E., Leng, J., Cotney, J., Reilly, S.K., Yin, J., Sarro, R., and Noonan, J.P. (2013). The genomic landscape of cohesin-associated chromatin interactions. *Genome Res.* 23, 1224–1234.
- Diep, D., Plongthongkum, N., Gore, A., Fung, H.-L., Shoemaker, R., and Zhang, K. (2012). Library-free methylation sequencing with bisulfite padlock probes. *Nat. Methods* 9, 270–272.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
- Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16, 1299–1309.
- Down, T.A., Rakan, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Gräf, S., Johnson, N., Herrero, J., Tomazou, E.M., et al. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* 26, 779–785.
- Egelhofer, T.A., Minoda, A., Klugman, S., Lee, K., Kolasinska-Zwier, P., Alekseyenko, A.A., Cheung, M.-S., Day, D.S., Gadel, S., Gorchakov, A.A., et al. (2011). An assessment of histone-modification antibody quality. *Nat. Struct. Mol. Biol.* 18, 91–93.
- ENCODE Project Consortium, Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
- Feinberg, A.P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature* 447, 433–440.
- Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. *Nature* 421, 448–453.
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473, 398–402.
- Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.
- Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet.* 10, 241–251.
- Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor- $\alpha$ -bound human chromatin interactome. *Nature* 462, 58–64.
- Gaffney, D.J., McVicker, G., Pai, A.A., Fondufe-Mittendorf, Y.N., Lewellen, N., Michelini, K., Widom, J., Gilad, Y., and Pritchard, J.K. (2012). Controls of nucleosome positioning in the human genome. *PLoS Genet.* 8, e1003036.

- Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405.
- Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. *Cell* 153, 17–37.
- Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature* 489, 91–100.
- Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* 153, 1149–1163.
- Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17, 877–885.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., and van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948–951.
- 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.
- Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* 43, 630–638.
- Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6, 479–491.
- Heintzman, N.D., and Ren, B. (2009). Finding distal regulatory elements in the human genome. *Curr. Opin. Genet. Dev.* 19, 541–549.
- 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.
- 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.
- Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science* 315, 1141–1143.
- Henikoff, S., Strahl, B.D., and Warburton, P.E. (2008). Epigenomics: a road-map to chromatin. *Science* 322, 853.
- Hindorf, L.A., and MacArthur, J. (European Bioinformatics Institute), Morales, J. (European Bioinformatics Institute), Junkins, H.A., Hall, P.N., Klemm, A.K., and Manolio, T.A. (2013). A Catalog of Published Genome-Wide Association Studies. Available at: [www.genome.gov/gwastudies](http://www.genome.gov/gwastudies).
- Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Biles, J.A., Birney, E., et al. (2013). Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* 41, 827–841.
- Holliday, R. (1990). Mechanisms for the control of gene activity during development. *Biol. Rev. Camb. Philos. Soc.* 65, 431–471.
- Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* Published online September 13, 2013. <http://dx.doi.org/10.1038/ng.2746>.
- Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R., and Rao, A. (2010). The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* 5, e8888.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303.
- Ji, H., Ehrlich, L.I.R., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467, 338–342.
- Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* 10, 161–172.
- Jin, S.-G., Kadam, S., and Pfeifer, G.P. (2010). Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* 38, e125.
- Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502.
- Jones, P.A., and Martienssen, R. (2005). A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop. *Cancer Res.* 65, 11241–11246.
- Kalhor, R., Tjong, H., Jayatilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* 30, 90–98.
- Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* 22, 2497–2506.
- 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.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Krueger, F., Kreck, B., Franke, A., and Andrews, S.R. (2012). DNA methylome analysis using short bisulfite sequencing data. *Nat. Methods* 9, 145–151.
- Laird, P.W. (2010). Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* 11, 191–203.
- Lander, E.S. (2011). Initial impact of the sequencing of the human genome. *Nature* 470, 187–197.
- 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.; International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831.
- Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. *Cell* 152, 1237–1251.
- Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.-I., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301–313.
- Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926.
- Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. *Cell* 128, 707–719.
- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.

- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536.
- Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322.
- Madhani, H.D., Francis, N.J., Kingston, R.E., Kornberg, R.D., Moazed, D., Narlikar, G.J., Panning, B., and Struhl, K. (2008). Epigenomics: a roadmap, but to where? *Science* 322, 43–44.
- Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190–1195.
- Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* 18, 1073–1083.
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877.
- Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr., Kinney, J.B., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 30, 271–277.
- 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.
- Mikkelsen, T.S., Xu, Z., Zhang, X., Wang, L., Gimble, J.M., Lander, E.S., and Rosen, E.D. (2010). Comparative epigenomic analysis of murine and human adipogenesis. *Cell* 143, 156–169.
- Murphy, P.J., Cipriany, B.R., Wallin, C.B., Ju, C.Y., Szeto, K., Hagarman, J.A., Benitez, J.J., Craighead, H.G., and Soloway, P.D. (2013). Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells. *Proc. Natl. Acad. Sci. USA* 110, 7772–7777.
- Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell* 154, 490–503.
- Neph, S., Stergachis, A.B., Reynolds, A., Sandstrom, R., Borenstein, E., and Stamatoiyannopoulos, J.A. (2012a). Circuitry and dynamics of human transcription factor regulatory networks. *Cell* 150, 1274–1286.
- Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012b). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489, 83–90.
- Ng, J.-H., Kumar, V., Muratani, M., Kraus, P., Yeo, J.-C., Yaw, L.-P., Xue, K., Lufkin, T., Prabhakar, S., and Ng, H.-H. (2013). In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev. Cell* 24, 324–333.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.
- Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257.
- Ong, C.-T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* 12, 283–293.
- Paige, S.L., Thomas, S., Stoick-Cooper, C.L., Wang, H., Maves, L., Sandstrom, R., Pabon, L., Reinecke, H., Pratt, G., Keller, G., et al. (2012). A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* 151, 221–232.
- Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473, 394–397.
- Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* 14, 341–356.
- Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* 30, 265–270.
- Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211.
- Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295.
- Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122, 517–527.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283.
- Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., and Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* 11, 633–648.
- Raiber, E.-A., Beraldi, D., Ficiz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* 13, R69.
- Rajagopal, N., Xie, W., Li, Y., Wagner, U., Wang, W., Stamatoiyannopoulos, J., Ernst, J., Kellis, M., and Ren, B. (2013). RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput. Biol.* 9, e1002968.
- Ram, O., Goren, A., Amit, I., Shores, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., et al. (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* 147, 1628–1639.
- Ramalingam, S., Annaluru, N., and Chandrasegaran, S. (2013). A CRISPR way to engineer the human genome. *Genome Biol.* 14, 107.
- Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147, 1408–1419.
- Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166.
- Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4, 651–657.
- Robertson, A.B., Dahl, J.A., Vågbo, C.B., Tripathi, P., Krokan, H.E., and Klungland, A. (2011). A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* 39, e55.
- Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., et al.; modENCODE Consortium. (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330, 1787–1797.
- Sajan, S.A., and Hawkins, R.D. (2012). Methods for identifying higher-order chromatin structure. *Annu. Rev. Genomics Hum. Genet.* 13, 59–82.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489, 109–113.



- 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.
- Serre, D., Lee, B.H., and Ting, A.H. (2010). MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* 38, 391–399.
- Sheffield, N.C., Thurman, R.E., Song, L., Safi, A., Stamatoyannopoulos, J.A., Lenhard, B., Crawford, G.E., and Furey, T.S. (2013). Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res.* 23, 777–788.
- Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.-L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153, 692–706.
- Smallwood, A., and Ren, B. (2013). Genome organization and long-range regulation of gene expression by enhancers. *Curr. Opin. Cell Biol.* 25, 387–394.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220.
- Song, C.-X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* 29, 68–72.
- Song, C.-X., Yi, C., and He, C. (2012). Mapping recently identified nucleotide variants in the genome and transcriptome. *Nat. Biotechnol.* 30, 1107–1116.
- Song, C.-X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153, 678–691.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495.
- Statham, A.L., Robinson, M.D., Song, J.Z., Coolen, M.W., Stirzaker, C., and Clark, S.J. (2012). Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res.* 22, 1120–1127.
- Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* 154, 888–903.
- Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., et al. (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* 7, e1002154.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1016–1028.
- The International Cancer Genome Consortium. (2010). International network of cancer genome projects. *Nature* 464, 993–998.
- Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. *Nature* 489, 75–82.
- Tian, Z., Tolić, N., Zhao, R., Moore, R.J., Hengel, S.M., Robinson, E.W., Stenoien, D.L., Wu, S., Smith, R.D., and Paša-Tolić, L. (2012). Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.* 13, R86.
- van Steensel, B., and Dekker, J. (2010). Genomics tools for unraveling chromosome architecture. *Nat. Biotechnol.* 28, 1089–1095.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351.
- Waddington, C.H. (1959). Canalization of development and genetic assimilation of acquired characters. *Nature* 183, 1654–1655.
- Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* 151, 206–220.
- Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918.
- Watson, J.D., and Crick, F.H. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171, 964–967.
- Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* 13, 36–47.
- Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.-Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A.C., Rappilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343–348.
- Williamson, I., Eskeland, R., Lettice, L.A., Hill, A.E., Boyle, S., Grimes, G.R., Hill, R.E., and Bickmore, W.A. (2012). Anterior-posterior differences in HoxD chromatin topology in limb development. *Development* 139, 3157–3167.
- Won, K.-J., Zhang, X., Wang, T., Ding, B., Raha, D., Snyder, M., Ren, B., and Wang, W. (2013). Comparative annotation of functional regions in the human genome using epigenomic data. *Nucleic Acids Res.* 41, 4423–4432.
- Xie, M., Hong, C., Zhang, B., Lowdon, R.F., Xing, X., Li, D., Zhou, X., Lee, H.J., Maire, C.L., Ligon, K.L., et al. (2013a). DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat. Genet.* 45, 836–841.
- Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013b). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153, 1134–1148.
- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149, 1368–1380.
- Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309, 626–630.
- Zhao, Z., Tavoosidana, G., Sjölander, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38, 1341–1347.
- Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152, 642–654.
- Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481.