



## Practice of Epidemiology

# Genome-Wide Epigenetic Studies in Human Disease: A Primer on -Omic Technologies

Huihuang Yan\*, Shulan Tian, Susan L. Slager, Zhifu Sun, and Tamas Ordog

\* Correspondence to Dr. Huihuang Yan, Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, 200 1st Street SW, Rochester, MN 55905 (e-mail: yan.huihuang@mayo.edu).

Initially submitted March 4, 2015; accepted for publication July 9, 2015.

Epigenetic information encoded in covalent modifications of DNA and histone proteins regulates fundamental biological processes through the action of chromatin regulators, transcription factors, and noncoding RNA species. Epigenetic plasticity enables an organism to respond to developmental and environmental signals without genetic changes. However, aberrant epigenetic control plays a key role in pathogenesis of disease. Normal epigenetic states could be disrupted by detrimental mutations and expression alteration of chromatin regulators or by environmental factors. In this primer, we briefly review the epigenetic basis of human disease and discuss how recent discoveries in this field could be translated into clinical diagnosis, prevention, and treatment. We introduce platforms for mapping genome-wide chromatin accessibility, nucleosome occupancy, DNA-binding proteins, and DNA methylation, primarily focusing on the integration of DNA methylation and chromatin immunoprecipitation–sequencing technologies into disease association studies. We highlight practical considerations in applying high-throughput epigenetic assays and formulating analytical strategies. Finally, we summarize current challenges in sample acquisition, experimental procedures, data analysis, and interpretation and make recommendations on further refinement in these areas. Incorporating epigenomic testing into the clinical research arsenal will greatly facilitate our understanding of the epigenetic basis of disease and help identify novel therapeutic targets.

cancer; complex diseases; DNA methylation; epigenome; histone modification

Abbreviations: ChIP-seq, chromatin immunoprecipitation–sequencing; CpG, cytosine-guanine dinucleotide; DMR, differentially methylated region; GWAS, genome-wide association studies; 5hmC, 5-hydroxymethylcytosine; MBD-seq, methyl-CpG-binding domain protein-based immunoprecipitation and sequencing; 5meC, 5-methylcytosine; MeDIP-seq, methylated DNA immunoprecipitation and sequencing; PCR, polymerase chain reaction; RRBS, reduced representation bisulfite sequencing; SNP, single-nucleotide polymorphism; WGBS, whole-genome bisulfite sequencing.

Epigenetics is the study of mitotically or meiotically heritable changes in gene expression that occur without changes in DNA sequence and are mediated by postsynthetic modifications of DNA and histone proteins, histone variants, noncoding RNAs, and the proteins that regulate and interpret the modifications (1, 2). The term “epigenome” refers to the genome-wide distribution of these marks and regulators (3–5). Acting “above and beyond” the normally static genome, hundreds of epigenomes define the unique transcriptional programs of our cells throughout development and aging (3, 6) under the influence of stochastic and environmental factors (1, 7, 8). Epigenetic mechanisms have been directly implicated

in disease susceptibility and progression (2, 3). The epigenome’s ability to translate environmental exposures into gene regulation also provides a link to disease.

The development of assays based on microarrays and next-generation sequencing has allowed the identification of mutations in chromatin regulators and other epigenetic variation related to human disease. Since epigenetic information varies not only between individuals but also between cell types and their functional states, accurate detection of epigenetic changes in patients requires the generation of a large number of reference epigenomes. The International Human Epigenome Consortium (IHEC) aims to generate over 1,000 reference

epigenomes within a decade (9). As part of the International Human Epigenome Consortium, the National Institutes of Health Roadmap Epigenomics Project and the National Human Genome Research Institute Encyclopedia of DNA Elements (ENCODE) Consortium have generated genome-wide maps for DNA methylation, histone modifications, transcription factor binding, and chromatin accessibility from over 200 cell or tissue types (10–12) (see Appendix). This joint effort has advanced our understanding of the interindividual variations and cell-type specificity of key epigenetic marks (13–15). Rapid development in epigenomic technology and the increasing body of genomic and epigenomic data offer an unprecedented opportunity to delineate how the interplay between genetic, environmental, and epigenetic components triggers complex diseases (16). In this review, we describe how high-throughput epigenomic technologies could be applied to population-based studies of diseases/traits. Focusing on DNA methylation and chromatin immunoprecipitation–sequencing (ChIP-seq) assays, we discuss challenges and possible solutions in experimental design, data analysis, and interpretation of genome-wide epigenomic studies.

## EPIGENETIC MARKS AND REGULATORS

### Chromatin structure and function

The nucleosome, the fundamental unit of chromatin, is composed of 147 base pairs of DNA wrapped around a histone octamer composed of 2 copies each of the 4 canonical histones or their variants (6, 17). DNA and histones are subject to covalent modifications, and there are at least 18 different functional groups (18). These modifications are dynamic and reversible, mediated by proteins that add (“writers”), interpret (“readers”), or remove (“erasers”) them (19, 20).

Several factors are responsible for shaping chromatin organization and DNA packaging (5, 21). ATP-dependent chromatin remodeling complexes regulate the spacing, positioning, deposition and ejection, and histone composition of nucleosomes (21–23). Regions occupied by densely packed nucleosomes block transcription factors from accessing the DNA, while nucleosome-free regions like these at the transcriptional start sites of actively transcribed genes are more “open” to transcription factor binding (24, 25). Several next-generation sequencing–based methods have been developed to map nucleosome positions and chromatin accessibility (5, 26) (Table 1). Furthermore, the chromatin fiber is folded into 3-dimensional structures, thus enabling physical interactions between gene regulatory elements that are often far away from each other on the linearized DNA but are regulated by the same mechanisms (27–29).

### DNA methylation

DNA methylation refers to the addition of a methyl group (me) at position 5 of the cytosine base (5meC), occurring mostly in cytosine-guanine dinucleotides (CpG’s) (30, 31). 5meC also occurs in non-CpG contexts in embryonic stem cells (32–34), adult neurons (35, 36), and some other types of cells (35, 36). DNA methylation is mediated by the DNA methyltransferase family. 5meC can be further modified by

ten-eleven translocation (TET) family dioxygenases, which sequentially oxidize 5meC to 5-hydroxymethylcytosine (5hmeC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC) (11, 37, 38). Additional reactions can eventually restore unmodified cytosine (39). The exact function of these oxidation derivatives remains unknown. It has been suggested that 5hmeC may represent an intermediate product in an active demethylation pathway (37, 40, 41). 5hmeC is highly enriched in neurons and may serve as an epigenetic mark involved in neural development (41).

In normal human cells, 5meC occurs at approximately 70%–80% of CpGs, mostly in low-density CpG regions (42, 43). The so-called CpG islands, which have a high CpG and high (>50%) GC content, are infrequently methylated (33, 43, 44). CpG islands represent approximately 2% of the human genome (45) and occupy approximately 60% of promoter regions (33). While the majority of gene promoters are devoid of DNA methylation, marked methylation is observed within actively transcribed gene bodies (9, 46).

DNA methylation, which is highly divergent across different organisms and cell types (11, 34), plays essential roles in mammalian development, X chromosome inactivation, transposable element repression, and genomic imprinting (11, 32, 34, 37). DNA methylation is influenced by genetic, stochastic, and environmental factors and is involved in developmental disorders and cancer (11, 42, 46–49).

### Histone modifications

The core histones (H2A, H2B, H3, and H4) each undergo a variety of modifications, primarily on their N-terminal “tails” (11, 50). The common modifications include lysine (K) mono-, di-, and trimethylation (Kme1/2/3), arginine (R) mono- and dimethylation (Rme1/2), and K acetylation (Kac) and ubiquitination, as well as serine/threonine/tyrosine phosphorylation (17, 18, 38, 51).

Histone modifications play fundamental roles in regulating gene transcription, DNA replication, and DNA repair by influencing chromatin states and serving as docking sites for other proteins like histone modification “readers” and “writers” (9, 38). Histone acetylation is generally associated with transcriptional activation, whereas methylation can activate or repress transcription, depending on the residue modified and the level of modification (9, 11, 52). Genome-wide profiling of histone modifications has revealed their association with distinct genomic regions such as enhancers, promoters, and gene bodies (11, 53) (Figure 1).

## EPIGENETIC MECHANISMS OF DISEASE

### Abnormal epigenetic profiles in disease

Epigenetic abnormalities have been detected in developmental syndromes, cancer (54), and nonmalignant diseases, including inflammatory (55, 56) and psychiatric (57) disorders. Epigenetic changes can occur locally, affecting individual genes, or globally, such as genome-wide DNA hypomethylation (33), loss of DNA methylation boundaries at CpG islands (58), and loss of balance between active and repressive histone modifications (59).

**Table 1.** High-Throughput Epigenomic Assays

Assay	Target(s)	Features	References
MNase-seq	Nucleosome occupancy and positioning Localization of DNA-binding proteins	Maps both histone and nonhistone proteins Needs high sequencing depth Bias toward AT-rich regions	180–182
DNase-seq	Chromatin accessibility	Maps cis-regulatory regions Generates base-pair resolution footprints for some transcription factors Requires high sequencing depth DNase I cleavage bias May miss some distal regulatory regions	73, 174, 183, 184
FAIRE-seq	Chromatin accessibility	Maps cis-regulatory regions Does not rely on any antibody or enzyme digestion Relatively lower signal enrichment compared with DNase-seq May miss some promoter regions	185–187
ATAC-seq	Chromatin accessibility	Maps nucleosome positioning, chromatin accessibility, and transcription factor binding sites simultaneously Localization of DNA-binding proteins Nucleosome positioning Requires fewer cells Fast protocol Primarily targets nucleosomes around regulatory regions	188, 189
ChIP-seq	Localization of DNA-binding proteins	Maps DNA-binding proteins Often requires large quantities of starting material Needs a high-quality antibody	10, 12, 14, 77, 142, 144, 190, 191
WGBS	DNA methylation	Covers approximately 95% of CpGs Accurate quantification Single-base resolution High proportion of uninformative reads from non-CpG regions Does not differentiate between 5meC and 5hmeC High cost	13, 15, 32, 96
RRBS	DNA methylation	Covers approximately 10%–20% of CpGs, largely in CpG islands Accurate quantification Single-base resolution Low cost Limited coverage Does not differentiate between 5meC and 5hmeC	32, 45, 192, 193

Table continues

Dysregulation of DNA methylation is pervasive in disease (43, 60, 61). Cancer cells display both global DNA hypomethylation and focal hypermethylation (33, 37, 62). The former often happens in low-density CpG regions (43), repeats, satellite DNA (33, 62), and regions associated with nuclear envelope lamina (43, 44). DNA hypomethylation can spread over large blocks of DNA up to a few megabases (58, 59). In contrast, locus-specific DNA hypermethylation occurs in CpG islands and lower-CpG-density CpG island

“shores” (within 2 kilobases of the CpG islands) (33, 43). Hypermethylation of CpG islands in promoters leads to silencing of genes, including tumor suppressors (37, 63, 64). Promoter hypomethylation can aberrantly activate oncogenes (33, 37, 64) and cause loss of imprinting (33, 47).

Global loss of active histone marks and global loss or gain of repressive marks have been identified across a variety of cancers. Histone acetylation and methylation are most commonly affected in cancer (51). Reduced binding has been

Table 1. Continued

Assay	Target(s)	Features	References
MeDIP-seq	DNA methylation	Covers approximately 60%–90% of CpGs Can distinguish between 5meC and 5hmeC Can target a large fraction of repeats regions Low cost Quantitative estimation Low (approximately 150-base-pair) resolution Difficult to identify small methylation changes Reduced sensitivity to less methylated regions	<a href="#">32</a> , <a href="#">123</a> , <a href="#">124</a> , <a href="#">139</a> , <a href="#">193</a> , <a href="#">194</a>
MBD-seq	DNA methylation	Covers approximately 60% of CpGs Can target a large fraction of repeats regions Low cost Quantitative estimation Low (approximately 150-base-pair) resolution Difficult to identify small methylation changes Reduced sensitivity to less methylated regions	<a href="#">32</a> , <a href="#">60</a> , <a href="#">126</a> , <a href="#">127</a>
27K array <sup>a</sup>	DNA methylation	Covers 27,578 CpGs Single-base resolution Low cost Limited coverage Does not differentiate between 5meC and 5hmeC	<a href="#">61</a> , <a href="#">79</a> , <a href="#">84</a> , <a href="#">104</a> , <a href="#">120</a> , <a href="#">121</a> , <a href="#">193</a>
450K array <sup>b</sup>	DNA methylation	Covers 482,421 CpGs Single-base resolution Low cost Ideal for screening large cohorts Limited coverage Does not estimate allele-specific methylation Does not differentiate between 5meC and 5hmeC	<a href="#">48</a> , <a href="#">91</a> , <a href="#">105</a> , <a href="#">122</a>

Abbreviations: A, adenine; ATAC-seq, assay for transposase-accessible chromatin using sequencing; C, cytosine; ChIP-seq, chromatin immunoprecipitation and sequencing; CpG, cytosine-guanine dinucleotide; DNase-seq, DNase I digestion and sequencing; FAIRE-seq, formaldehyde-assisted isolation of regulatory elements sequencing; G, guanine; 5hmeC, 5-hydroxymethylcytosine; MBD-seq, methyl-CpG-binding domain protein-based capture and sequencing; 5meC, 5-methylcytosine; MeDIP-seq, methylated DNA immunoprecipitation and sequencing; MNase-seq, micrococcal nuclease digestion and sequencing; RRBS, reduced representation bisulfite sequencing; T, thymine; WGBS, whole-genome bisulfite sequencing.

<sup>a</sup> Illumina Infinium HumanMethylation27 BeadChip array (Illumina, Inc., San Diego, California).

<sup>b</sup> Illumina Infinium HumanMethylation450 BeadChip array (Illumina, Inc.).

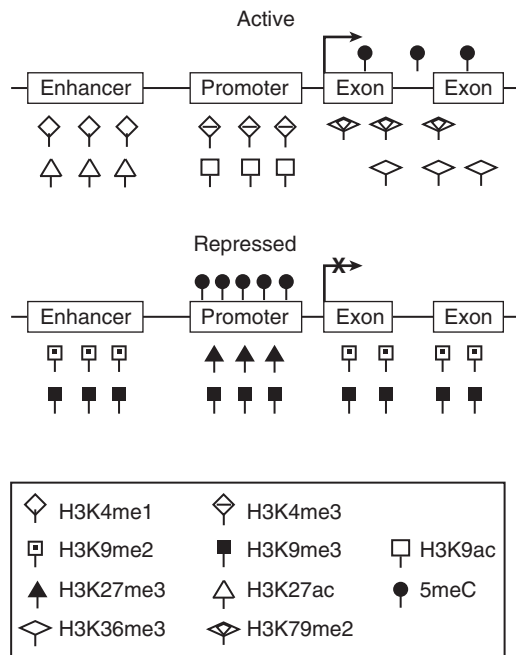
reported for the active marks H3K9ac (acetylated lysine 9 of histone H3), H3K18ac, H3K4me1/2/3, H4K5ac, H4K8ac, and H4K16ac, as well as the repressive marks H4K20me3 and H4R3me2 (9, 33, 65). The repressive marks H3K27me3 and H3K9me2/3 show either increased or decreased binding (33, 65, 66).

### Mechanisms of epigenetic alterations in disease

Multiple mechanisms are involved in the disruption of the epigenome in disease (54, 67). Altered cellular signaling can change the epigenome through the misregulation of epigenetic writers, readers, and erasers. Chromatin regulators, which often play roles in cancer as oncogenes or tumor suppressors (68), are frequently mutated (63, 69), translocated (65), or

aberrantly expressed (33, 59) in various diseases. Mutations of transcription factors that are often cancer drivers (68) may also contribute to tumorigenesis (69). This topic has been extensively reviewed (19, 37, 59, 69–72).

Recent studies have suggested roles for epigenetic mechanisms in linking DNA variants to disease phenotypes (50). Genome-wide association studies (GWAS) have identified single-nucleotide polymorphisms (SNPs) associated with many complex human diseases or traits. Approximately 90% of the SNPs reside in noncoding regions, particularly in regions sensitive to DNase I cleavage, suggesting regulatory roles for SNPs in gene expression (50, 73). Specifically, variants lying within canonical binding motifs can directly affect transcription factor occupancy by disrupting these binding sites or creating new ones (52, 74–77). Alteration of transcription



**Figure 1.** Epigenetic features that mark active and repressed genes. Active enhancers, promoters, and gene bodies are marked by H3K4me1+H3K27ac, H3K4me3+H3K9ac, and H3K79me2+H3K36me3, respectively; 5-methylcytosine (5mC) is often detected in the gene bodies. Inactive enhancers, promoters, and gene bodies are marked by H3K9me2/3 and H3K27me3; 5mC also occurs in the promoters. H3K4me1, histone H3 lysine 4 monomethylation; H3K4me3, histone H3 lysine 4 trimethylation; H3K9ac, histone H3 lysine 9 acetylation; H3K9me2, histone H3 lysine 9 dimethylation; H3K9me3, histone H3 lysine 9 trimethylation; H3K27ac, histone H3 lysine 27 acetylation; H3K27me3, histone H3 lysine 27 trimethylation; H3K36me3, histone H3 lysine 36 trimethylation; H3K79me2, histone H3 lysine 79 dimethylation.

factor binding affinity may lead to changes in gene expression (76), possibly through the recruitment of histone-modifying enzymes (i.e., enzymes which catalyze the addition or removal of histone modifications) that determine local chromatin states (50, 52). Genetic variants can influence multiple layers of epigenomic information. For example, in lymphoblastoid cell lines, a single genetic variant has been shown to alter nucleosome position, chromatin accessibility, histone modifications, and RNA polymerase II occupancy (52). Thus, the interaction of genetic variation and epigenetic mechanisms increases susceptibility to disease (69, 72).

Aging is a key risk factor for many diseases (78). Some environmental and lifestyle factors, such as smoking (7, 55, 79), alcohol (80), diet (55), and life stress (37), have also been linked to aging-associated diseases. The epigenome is an emerging link between aging and disease and between environmental factors and disease (37, 47, 81). The plasticity of the epigenome enables an organism to respond to both external and internal stimuli (7, 9). The organism can accumulate these experiences over time, and such an “epigenetic memory” may be maintained for life (9, 82) or even transmitted to the next generation (37, 83).

Both aging (37, 84) and environmental factors (7, 55, 82) can alter DNA methylation and/or histone modifications. Aging is accompanied by DNA hypermethylation of gene promoters and global hypomethylation (16, 62), which may in part reflect the accumulation of environmental exposures over time (81). Ethanol affects DNA methylation (85) and/or histone modifications (80, 85), as do smoking (79, 86, 87) and nutritional factors (55, 88). The contribution of environmental exposures and lifestyles to epigenetic variation is being intensively investigated in large cohorts (2).

## GENOME-WIDE DNA METHYLATION ASSAYS

Early studies inferred overall methylation status by digesting genomic DNA samples in parallel with both methylation-sensitive restriction enzymes (e.g., *HpaII*) and their methylation-insensitive isoschizomers (e.g., *MspI*). Later, sodium bisulfite, which converts unmethylated but not methylated or hydroxymethylated cytosines into uracil, was introduced (11). Here, the introduced sequence differences from bisulfite treatment allow the assessment of methylation status at individual cytosines (31, 89). Large-scale DNA methylation profiling is performed by microarray hybridization, affinity-based assays followed by sequencing, or sequencing of bisulfite-converted DNA (34, 81, 90–92) (Table 1). These approaches differ in cost, resolution, accuracy, genomic coverage, and specificity for 5mC versus 5hmC (2). 5hmC and 5-formylcytosine can be selectively assessed by means of affinity- and chemical modification-based approaches coupled with next-generation sequencing (93–95). Since these methods have not been utilized in epidemiologic studies, they are not discussed further here.

## Bisulfite sequencing platforms

Whole-genome bisulfite sequencing (WGBS) aims to quantify the methylation status of each cytosine in both CpG and non-CpG contexts (13). It is considered the “gold standard” for methylation profiling. However, the required high coverage (approximately 30× coverage) and associated high cost have limited its application (2, 34). In addition, 70%–80% of WGBS reads are not informative because they provide little information about CpG methylation (96). These reads are mostly mapped to non-CpG regions, with a small proportion mapped to “static” regions showing no changes or small changes in CpG methylation between different tissues or cell types (96). Reduced representation bisulfite sequencing (RRBS) is much more cost-effective, because it selectively targets CpG-rich sites. In RRBS, genomic DNA is digested with a methylation-insensitive restriction enzyme (e.g., *BglIII*, *MspI*) to enrich for genomic regions containing CpGs. The fragments are then subjected to bisulfite treatment. Thus, RRBS is biased toward CpG-rich sites such as CpG islands (9). In clinical settings, RRBS is very useful because it allows rapid screening of many patient samples in a cost-effective manner (9).

Bioinformatic analysis begins with checking sequence quality using software packages like *FastQC* (Babraham Institute, Cambridge, United Kingdom; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adaptor sequences attached during library preparation for sequencing can be removed



using the *cutadapt* program (Department of Computer Science, TU Dortmund University, Dortmund, Germany; <http://code.google.com/p/cutadapt/>). Two alignment strategies, “wild-card” and “3-letter,” are available for mapping reads to the reference genome. The wild-card approach, such as *BSMAP* (97, 98), which uses the Short Oligonucleotide Alignment Program, replaces cytosines in the reference sequence with the wild-card letter Y, which is defined to match both cytosines (i.e., originally (hydroxy)methylated and, therefore, unconverted cytosines) and thymines (i.e., originally unmethylated cytosines, converted to uracils by bisulfite treatment and then to thymines during polymerase chain reaction (PCR) amplification in library preparation) in the reads (42). In the 3-letter approach, all cytosines are first converted into thymines and guanines to adenines in both the reads and the reference sequence; mapping is then performed using a standard aligner like *Bowtie* (99) (implemented in *Bismark* (100)). These alignment methods achieve a similar coverage of CpG sites (101).

To quantify the methylation level of individual CpGs and non-CpGs at a given base C on the plus strand, the total number of C-carrying and T-carrying reads is counted and the methylation ratio is estimated as  $C/(C+T)$ . Similarly, for a given base G on the minus strand, the methylation ratio is estimated as  $G/(G+A)$  (102). The methylation level can also be calculated by summing up reads mapping to both strands (32). Several pipelines have been developed for streamlined analysis of WGBS and/or RRBS data (102, 103).

### Hybridization-based platforms

The Illumina Infinium HumanMethylation27 (27K) (104) and HumanMethylation450 (450K) BeadChip arrays (Illumina, Inc., San Diego, California) (105) have been most widely used in epigenome-wide association studies, largely because of their relatively low cost and high reproducibility (11, 89, 91, 106). The methylation module of GenomeStudio software (Illumina, Inc.) provides options for signal normalization and background subtraction using control probes (42). DNA methylation is assessed by calculating the fraction of methylated cytosines ( $\beta$  values) at individual CpGs (89, 107). Several software packages have been developed for analysis of the array data (108), including *lumi* for both the 27K and 450K arrays (109) and *minfi* for the 450K array (110).

### Affinity-based platforms

Methylated DNA immunoprecipitation and sequencing (MeDIP-seq) and methyl-CpG-binding domain protein-based capture and sequencing (MBD-seq) are cost-effective assays for genome-wide DNA methylation profiling, particularly for repetitive DNA regions (32, 111). In MeDIP-seq, genomic DNA is fragmented by sonication and an adaptor is ligated. The sample is denatured and immunoprecipitated with an anti-5meC antibody. The immunoprecipitated products are PCR-amplified, size-selected, and sequenced to a depth of about 30–50 million mapped reads (34, 45). DNA hydroxymethylation can be similarly assessed with an anti-5hmC antibody (112).

MBD-seq is performed similarly but without denaturing (32). It uses recombinant methyl-CpG-binding domain from

methyl-CpG-binding domain protein 2 (MBD2 or MeCP2) as the primary affinity reagent (60). In both methods, the local enrichment level is highly correlated with CpG density (113). MBD-seq is more effective in identifying methylated regions containing multiple methylated cytosines; in contrast, MeDIP-seq often recovers regions with sporadically methylated CpGs of presumably less biological relevance (60). Affinity enrichment methods only provide qualitative estimation of DNA methylation (9) and have low resolution (11, 34, 92). They are also less sensitive to methylated regions with low CpG density than bisulfite sequencing (11, 34) and are less powerful in detecting small changes (11).

Publicly available software packages for analyzing MeDIP-seq and MBD-seq data include *Batman* (114), *MEDIPS* (115), and *BayMeth* (part of the *Repitools* package) (113). For example, *MEDIPS* takes mapping output and assigns reads to predefined regions of interest or fixed-size windows. It checks overall enrichment and saturation, estimates relative methylation score after CpG density normalization using linear modeling, identifies differentially methylated regions (DMRs) using *edgeR*, and annotates methylated regions relative to genomic features (115).

### Differential methylation analysis

Differential methylation analysis seeks to determine significant differences in methylation levels between groups of samples (107). It can be performed on individual CpGs or, for increased statistical power, on predefined regions (e.g., CpG islands, CpG island shores) or sliding windows (42). Software packages for identifying differentially methylated cytosines or DMRs from the above 3 platforms have been reviewed (42, 107, 108, 116).

For WGBS and RRBS data analysis, the *methylKit* package uses Fisher’s exact test in the absence of biological replicates and logistic regression ( $\beta$  binomial model) when biological replicates are available (117). The *BSmooth* pipeline is tailored to the analysis of low-coverage WGBS data, including the function for identifying DMRs (118). Several packages provide functions to identify differentially methylated cytosines, differentially methylated positions, or DMRs from methylation arrays (107, 108, 116). The Wilcoxon rank-sum test (119), moderated *t* test (116), and *F* test (110, 116) are commonly applied to normalized log ratios of intensities. The *IMA* (119) and *minfi* (110) packages allow both site- and region-level analysis of differential methylation. For MeDIP-seq and MBD-seq data, DMRs can be identified using *edgeR* implemented in the *MEDIPS* pipeline (115).

Tissue-specific or disease-related DMRs are largely located at distal regulatory regions, CpG island shores, or low-CpG-density promoters rather than at promoter CpG islands (11, 59). It is important to verify and validate the identified DMRs (92). A simple way is to check some of the DMRs in a genome browser and look for signs of artifacts. A variety of quality control plots provide global views about the magnitude of the changes and genomic distribution of DMRs (42). In addition, it is indispensable to experimentally verify the reproducibility of the DMRs using a different assay and validate the findings in a different cohort (2, 42).

## DNA methylation case studies

The Illumina Infinium BeadArrays have been the method of choice in epidemiologic studies (48, 84, 120–122). More recently, MeDIP-seq (123–125) and MBD-seq (126, 127) have begun to gain popularity, owing to their much wider genomic coverage and relatively lower cost.

Traditional epidemiologic study designs, like case-control and cohort studies, have been applied to epigenetic studies (2, 47). These designs need to include unrelated, age- and sex-matched normal controls (88) or case-control individuals (124). Monozygotic twins are genetically identical and thus provide a powerful approach for assessing the role of the epigenome in mediating environmental and lifestyle risk factors involved in complex diseases (7, 24, 37, 47, 92). Several ongoing large-scale cohort studies, such as Accessible Resource for Integrated Epigenomics Studies (ARIES), which uses a birth cohort (<http://www.ariesepigenomics.org.uk/>), and EpiTwin, which uses a twin cohort (<http://www.epitwin.eu/index.html>), aim to correlate methylation changes with environmental exposures and disease development (2, 125, 128).

These studies have identified DMRs or differentially methylated positions associated with complex diseases, including autoimmune diseases (88, 121), diabetes (123, 125, 129), heat pain (123), major depressive disorder (124), autism spectrum disorder (130), schizophrenia (126), and rheumatoid arthritis (122). Significant changes in DNA methylation also occur in response to environmental factors (2, 55, 131, 132) and during aging (48, 84, 120, 127, 133).

## Challenges in DNA methylation studies

Despite their widespread use in epidemiology, DNA methylation studies face several challenges, from sample collection to data analysis and interpretation. Firstly, DNA methylation analysis often uses whole tissue containing several different cell types, each with its own unique DNA methylome (56, 134). Thus, some of the DMRs may simply reflect the differences in cell composition (42, 47). Indeed, much of the variability in DNA methylation previously thought to be associated with aging is actually caused by aging-dependent changes in cell content (49). In addition, true methylation differences may be averaged out if they exist only in a particular type of cell.

Algorithms have been developed for *in silico* deconvolution of array-based methylation signals in whole blood (49, 135), taking advantage of prior knowledge of the dominant cell types' methylation profiles (116). Briefly, cell-type-specific differentially methylated cytosines were first identified using 5–6 dominant cell populations generated by cell sorting (49, 135). The top differentially methylated cytosines were then used to build a regression model, which was finally used to estimate the relative proportions of these major cell types in new blood samples based on existing methylation profiles. This method, available in the *minfi* package (110), was used to adjust for cell mixture in epigenome-wide association studies (122). Recently, 2 packages were developed to correct for difference in cell composition without relying on DNA methylation profiles from the major cell types (136, 137). In addition, several experimental methods are available for reducing cellular heterogeneity in a tissue, such as microdissection,

laser-capture microdissection, and fluorescence-activated cell sorting (2, 138). The applicability of these methods, however, is tissue-dependent.

Secondly, there are extensive variations in DNA methylation between different tissues—in fact, DNA methylation varies more between brain and blood from the same individual than between individuals in the same tissue (139). Therefore, in most cases it is critical to use the tissue most relevant to the disease under investigation (7, 30). Nevertheless, DNA methylation studies frequently use easily accessible surrogate tissues like whole blood (49, 56, 83, 91, 106). Although this practice remains controversial (55, 92, 140), it may well be justified if systemic involvement or immune/inflammatory etiology is suspected (141).

Thirdly, despite the development of platform-specific software packages, analyzing genome-wide DNA methylation data from large cohorts remains a challenge (2, 42, 116). In order to reduce artifacts and increase sensitivity, further efforts are needed to systematically benchmark existing tools and to develop more powerful analytical systems. Finally, since DNA methylation differences between cases and controls are usually quite small, especially in surrogate tissues (2), the findings need to be interpreted with caution.

## CHROMATIN IMMUNOPRECIPITATION-SEQUENCING

ChIP-seq is a powerful method for identifying global protein-DNA interactions (142, 143), including transcription factor binding (5), histone modifications (144), chromatin remodeling complex subunits (145), and RNA polymerase occupancy (146). In this assay, chromatin is cross-linked by formaldehyde and sheared into approximately 150- to 300-base-pair fragments by sonication and/or micrococcal nuclease digestion. The size distribution of chromatin fragments should be verified by an appropriate method. Genomic regions of interest are enriched using a factor-specific antibody. Purified ChIP DNA is then subjected to adaptor ligation, PCR amplification, and sequencing (11, 143, 147). A control library—for example, input or ChIP obtained with a nonspecific antibody—needs to be generated and sequenced in parallel (148).

## ChIP-seq experimental design

The outcome of a ChIP-seq experiment strongly depends on the optimization of a few key factors, such as antibody, amount of input material, tissue source and quality, and sequencing depth. Antibody quality is a key factor responsible for the reliability of the binding profile. Over 200 commercial antibodies against transcription factors, histone modifications, and chromatin regulators have been evaluated (148–151). While these public resources contain information about previously validated antibodies, assessing antibody quality before a ChIP experiment is highly recommended, since lot-to-lot variability is quite common.

If possible, a ChIP-seq experiment should include 2 biological replicates in order to assess reproducibility (26, 148). Most ChIP and control libraries are single-end sequenced to a length of approximately 50 base pairs. A minimum of 20 million uniquely mapped reads (10 million per replicate) is required for transcription factors and highly localized chromatin

marks (148), whereas 40 million reads are needed for assessing marks with diffuse binding profiles (148, 152). Depending on mark abundance, conventional ChIP-seq experiments typically require 1–10 million cells (143, 153). A few new protocols have used approximately 10,000 cells in ChIP-seq targeting histone modifications (147, 154) and transcription factors (154). Notable examples are the indexing-first chromatin immunoprecipitation (iChIP) approach (154) and the high-throughput chromatin immunoprecipitation (HT-ChIP) method (147); the latter allows automatic and sensitive ChIP-seq in a 96-well format.

### ChIP-seq data analysis

To identify binding sites, reads are first mapped back to the reference genome using a short-read aligner like *BWA* (155). Very often only uniquely mapped reads are kept and duplicates are filtered out. A peak caller is then used to identify binding sites and generate files for data visualization. Many packages have been developed for identifying peaks from discrete binding profiles, including the widely used model-based algorithm *MACS* (156). However, it remains challenging to define the boundary of broad peaks due to the discontinuity in signal distribution. *SICER*, which employs spatial clustering to address this issue, models read counts using the Poisson distribution (157). It scans for individual enriched windows, links nearby enriched windows (often allowing a gap of 3–5 windows) into a single domain, and then calculates statistics for domain enrichment over the control library. *RSEG*, which uses negative binomial distribution and corrects for mappability bias, shows performance comparable to that of *SICER* (158). Overall ChIP-seq quality can be assessed based on the quality metrics proposed by the Encyclopedia of DNA Elements Consortium (148, 159).

The resolution of ChIP-seq is limited by the size of sequenced fragments, often a few hundred base pairs. Generating high-resolution maps is important for defining DNA motifs preferred by sequence-specific transcription factors and other regulatory elements (5). Protein-DNA interactions can be mapped at base-pair resolution using ChIP-exo, which employs lambda exonuclease to digest immunoprecipitated chromatin not protected by protein-DNA crosslinks (160).

### Differential binding analysis

A number of approaches have been developed to identify differential binding sites between groups of samples (161–164). The *DiffBind* (162) and *DBChIP* packages (161) start with peaks, while *diffReps* (163) uses 1-kilobase sliding windows. For the former, peaks from all libraries need to be consolidated into a single peak list (161, 162). The number of reads overlapping each window or merged peak is then tabulated for individual ChIP and input libraries and normalized. Several methods have been proposed for ChIP-seq data normalization, including global scaling based on total mapped reads or number of reads within peaks (26), linear modeling based on  $M$  (log ratio) and  $A$  (mean on a log scale) values of shared peaks (165), quantile normalization (166), and trimmed mean of  $M$  values (167). To choose an appropriate normalization method, it is better to first check whether global

enrichment levels are comparable among different libraries. Recently, “spike-in” chromatin from a different organism was used as an internal control to facilitate data normalization across different samples (168, 169).

*DiffBind* can only be applied to data sets with biological replicates (162). It uses *edgeR* or *DESeq* to identify differential binding sites. *DBChIP* uses *edgeR* to estimate the dispersion parameter under the negative binomial distribution (161) and can also handle cases without biological replicates. In contrast, the sliding window-based *diffReps* selects significant windows that meet a predefined  $P$ -value cutoff, and then repeats the test of significance after merging neighboring significant windows into regions (163). An exact negative binomial test is used if there are biological replicates; otherwise, a  $G$  test or  $\chi^2$  test is used.

### ChIP-seq applications

ChIP-seq has been applied to a few case-control studies with a limited number of subjects (20 or less). These studies aimed to identify marks in regulatory regions that were associated with aging (134) or cancer development (162, 170). H3K4me1 and H3K27ac together mark active enhancers, and H3K4me1 alone marks poised enhancers. Compared with normal colon crypts, colorectal cancer is characterized by both gain of H3H4me1 sites in the open chromatin regions and loss of H3H4me1 sites from condensed chromatin regions, suggesting distinct chromatin changes in enhancer regions (170). Estrogen receptor  $\alpha$  is a key transcription factor in the majority of breast cancers. ChIP-seq revealed distinct estrogen receptor  $\alpha$  binding signatures that can separate patients with good clinical outcomes from those with poor outcomes, and the latter showed a global increase in estrogen receptor  $\alpha$  binding (162).

ChIP-seq is heavily used in post-GWAS functional studies by mapping transcription factor binding and histone modifications to regulatory regions, thereby allowing the inference of causality for noncoding SNPs (171). This area of research is rapidly expanding, due in part to the massive amounts of epigenomic data generated by the International Human Epigenome Consortium (14, 29, 144) and the recognition that noncoding SNPs play regulatory roles (52, 75, 77). In fact, a significant portion of causal SNP candidates overlap enhancers marked by H3K4me1 and H3K27ac (73, 77).

There are frequent cross-talks between different marks (33, 44, 52, 63). For example, in one study, DNA methylation and H3K27me3 together contributed to the gene silencing in hindbrain ependymomas (172). Whole-genome or exome sequencing failed to identify significant recurrent mutations across a cohort of 47 hindbrain ependymoma patients; however, array hybridization- and WGBS-based DNA methylation assays identified DNA methylation target genes that showed significant overlap with H3K27me3 targets in the group A subtype (with poor prognosis), a pattern that was much less obvious in the group B subtype (with good prognosis) (172).

Therefore, the epigenetic basis of disease can be more readily revealed through integrative analysis of genomic and epigenomic data sets (29, 173). Public data generated from the same or relevant cell/tissue types represent good resources (12, 73, 144, 174). Correlating expression data with epigenomic data like transcription factor or histone modification



ChIP-seq or DNA methylation will help identify the genes potentially targeted by an epigenetic mark (172). Furthermore, chromatin interaction maps generated by the chromosome conformation capture (3C)-based assays can be utilized to link gene promoters to the distal regulatory regions like enhancers (27–29). The combined analysis of variant calls with expression data, WGBS/RRBS data, DNase I hypersensitive sites, or ChIP-seq data can identify allele-specific events, thus pinpointing the disease-associated noncoding functional variants in linkage disequilibrium (14, 52, 77).

### Challenges in ChIP-seq experiments

Unlike DNA methylation experiments, which use genomic DNA and generally work well in archived samples (2, 175, 176), ChIP-seq experiments require fresh or frozen samples containing high-quality chromatin. In addition, the conventional ChIP protocol often needs large amounts of starting material ( $10^6$  cells or more) for a single mark, which limits its applicability to scarce clinical samples. In these cases, the indexing-first chromatin immunoprecipitation approach, which has high sensitivity and reproducibility, may become the method of choice (154).

Similarly to DNA methylation studies, the interpretation of ChIP-seq experiments performed on unfractionated tissues is complicated by the presence of multiple cell types. The ideal solution would be to use sorted cell populations (77, 154). Nevertheless, it is often difficult to obtain large numbers of homogeneous cells from cell sorting (140), and cell sorting itself might alter epigenetic states (3). Cell culturing also promotes epigenetic changes (3, 10), limiting its usefulness in enriching target cells (50, 81, 138). Lastly, given the variation in ChIP efficiency in different experiments (26, 159), it is difficult to directly compare data sets that were generated with antibodies from different sources, in different laboratories, or at different times. Further automation of the ChIP protocol is expected to reduce variability.

### CONCLUSIONS AND OUTLOOK

Epigenetic changes very likely contribute to the pathogenesis of complex diseases by mediating gene-environment interactions. Epigenome-wide association studies and post-GWAS functional analyses support this notion by revealing disease-associated changes in DNA methylation, histone modifications, transcription factor binding, and noncoding RNA expression. However, development of epigenomics-based therapeutic strategies requires establishment of causality (55). Several hurdles need to be cleared to achieve this goal.

Firstly, given the strong tissue specificity and highly dynamic nature of the epigenome (2, 9, 37, 177), it is essential to generate reference epigenomes for additional tissue types and developmental stages (9, 12, 16). Cataloging epigenome variations across normal individuals will enhance the identification of site-specific variations associated with disease (2, 12, 57) (Appendix). Establishment of concordance between easily accessible tissues and inaccessible ones of greater disease relevance (2) would facilitate not only biomarker discovery (124) but also longitudinal evaluation of epigenetic

changes in prospective cohort studies, which, in turn, should help differentiate cause from simple association (7, 56, 106, 175). These efforts should be accompanied by the development of more sensitive assay protocols (3), because current methods often require large quantities of tissues (38).

Analysis of large data sets represents another challenge. The Encyclopedia of DNA Elements Consortium has developed analytical guidelines for multiple types of epigenetic data (178). However, bioinformatic methods must also take into account the fact that technologies are frequently updated or replaced, and must offer solutions to improve compatibility across data generated over time or on different platforms (5, 16, 81). Additional tools for the integrative analysis of “multi-omic” data sets must be developed to enhance the dissection of disease mechanisms.

It is anticipated that progress in our understanding of disease epigenetics will lead to new disease diagnosis, prevention, and treatment (9, 61, 179). While the current exponential growth in this area of biomedical research supports this view, the path toward achieving this ambitious goal remains challenging.

### ACKNOWLEDGMENTS

Author affiliations: Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota (Huihuang Yan, Shulan Tian, Susan L. Slager, Zhifu Sun); Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota (Huihuang Yan, Zhifu Sun, Tamas Ordog); Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, Minnesota (Tamas Ordog); and Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minnesota (Tamas Ordog).

This study was supported by the Center for Individualized Medicine at the Mayo Clinic and by the National Institutes of Health (grants R01 DK058185, P01 DK068055, and R21 CA191186 to T.O.).

Conflict of interest: none declared.

### REFERENCES

1. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat Rev Genet.* 2007;8(4):253–262.
2. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet.* 2013;14(8):585–594.
3. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol.* 2010;28(10):1045–1048.
4. Satterlee JS, Schübeler D, Ng HH. Tackling the epigenome: challenges and opportunities for collaboration. *Nat Biotechnol.* 2010;28(10):1039–1044.
5. Zentner GE, Henikoff S. High-resolution digital profiling of the epigenome. *Nat Rev Genet.* 2014;15(12):814–827.
6. Sarda S, Hannonhalli S. Next-generation sequencing and epigenomics research: a hammer in search of nails. *Genomics Inform.* 2014;12(1):2–11.
7. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature.* 2010;465(7299):721–727.

8. D'Urso A, Brickner JH. Mechanisms of epigenetic memory. *Trends Genet.* 2014;30(6):230–236.
9. Maunakea AK, Chepelev I, Zhao K. Epigenome mapping in normal and disease states. *Circ Res.* 2010;107(3):327–339.
10. Zhu J, Adli M, Zou JY, et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell.* 2013;152(3):642–654.
11. Shin J, Ming GL, Song H. Decoding neural transcriptomes and epigenomes via high-throughput sequencing. *Nat Neurosci.* 2014;17(11):1463–1475.
12. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–330.
13. Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009;462(7271):315–322.
14. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011;473(7345):43–49.
15. Lister R, Mukamel EA, Nery JR, et al. Global epigenomic reconfiguration during mammalian brain development. *Science.* 2013;341(6146):1237905.
16. Almouzni G, Altucci L, Amati B, et al. Relationship between genome and epigenome—challenges and requirements for future research. *BMC Genomics.* 2014;15:487.
17. Henikoff S, Shilatifard A. Histone modification: cause or cog? *Trends Genet.* 2011;27(10):389–396.
18. Huang H, Sabari BR, Garcia BA, et al. SnapShot: histone modifications. *Cell.* 2014;159(2):458–458.e451.
19. Chi P, Allis CD, Wang GG. Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nat Rev Cancer.* 2010;10(7):457–469.
20. Marsh DJ, Shah JS, Cole AJ. Histones and their modifications in ovarian cancer—drivers of disease and therapeutic targets. *Front Oncol.* 2014;4:144.
21. Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta.* 2014;1839(8):627–643.
22. Bell O, Tiwari VK, Thomä NH, et al. Determinants and dynamics of genome accessibility. *Nat Rev Genet.* 2011;12(8):554–564.
23. Chen T, Dent SY. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet.* 2014;15(2):93–106.
24. Milosavljevic A. Emerging patterns of epigenomic variation. *Trends Genet.* 2011;27(6):242–250.
25. Struhl K, Segal E. Determinants of nucleosome positioning. *Nat Struct Mol Biol.* 2013;20(3):267–273.
26. Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat Rev Genet.* 2014;15(11):709–721.
27. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 2012;26(1):11–24.
28. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet.* 2013;14(6):390–403.
29. Rivera CM, Ren B. Mapping human epigenomes. *Cell.* 2013;155(1):39–55.
30. Birney E. Chromatin and heritability: how epigenetic studies can complement genetic approaches. *Trends Genet.* 2011;27(5):172–176.
31. Szyf M. DNA methylation signatures for breast cancer classification and prognosis. *Genome Med.* 2012;4(3):26.
32. Harris RA, Wang T, Coarfa C, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol.* 2010;28(10):1097–1105.
33. Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol.* 2010;28(10):1057–1068.
34. Li D, Zhang B, Xing X, et al. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. *Methods.* 2015;72:29–40.
35. Pinney SE. Mammalian non-CpG methylation: stem cells and beyond. *Biology (Basel).* 2014;3(4):739–751.
36. Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell.* 2014;156(1-2):45–68.
37. Brookes E, Shi Y. Diverse epigenetic mechanisms of human disease. *Annu Rev Genet.* 2014;48:237–268.
38. Hyun BR, McElwee JL, Soloway PD. Single molecule and single cell epigenomics. *Methods.* 2015;72:41–50.
39. Williams K, Christensen J, Helin K. DNA methylation: TET proteins—guardians of CpG islands? *EMBO Rep.* 2012;13(1):28–35.
40. Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nat Rev Genet.* 2014;15(10):647–661.
41. Kinde B, Gabel HW, Gilbert CS, et al. Reading the unique DNA methylation landscape of the brain: non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A.* 2015;112(22):6800–6806.
42. Bock C. Analysing and interpreting DNA methylation data. *Nat Rev Genet.* 2012;13(10):705–719.
43. Weisenberger DJ. Characterizing DNA methylation alterations from the Cancer Genome Atlas. *J Clin Invest.* 2014;124(1):17–23.
44. Bergman Y, Cedar H. DNA methylation dynamics in health and disease. *Nat Struct Mol Biol.* 2013;20(3):274–281.
45. Gu H, Smith ZD, Bock C, et al. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc.* 2011;6(4):468–481.
46. Yang X, Han H, De Carvalho DD, et al. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell.* 2014;26(4):577–590.
47. Rakyan VK, Down TA, Balding DJ, et al. Epigenome-wide association studies for common human diseases. *Nat Rev Genet.* 2011;12(8):529–541.
48. Martino D, Loke YJ, Gordon L, et al. Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol.* 2013;14(5):R42.
49. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 2014;15(2):R31.
50. Furey TS, Sethupathy P. Genetics driving epigenetics. *Science.* 2013;342(6159):705–706.
51. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell.* 2012;150(1):12–27.
52. McVicker G, van de Geijn B, Degner JF, et al. Identification of genetic variants that affect histone modifications in human cells. *Science.* 2013;342(6159):747–749.
53. Jakovcevski M, Akbarian S. Epigenetic mechanisms in neurological disease. *Nat Med.* 2012;18(8):1194–1204.
54. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet.* 2007;8(4):286–298.
55. Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med.* 2010;7(10):e1000356.

56. Murphy TM, Mill J. Epigenetics in health and disease: heralding the EWAS era. *Lancet*. 2014;383(9933):1952–1954.
57. Lelièvre SA. Taking a chance on epigenetics. *Front Genet*. 2014;5:205.
58. Hansen KD, Timp W, Bravo HC, et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*. 2011;43(8):768–775.
59. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat Rev Cancer*. 2013;13(7):497–510.
60. Serre D, Lee BH, Ting AH. MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res*. 2010;38(2):391–399.
61. Xu Z, Bolick SC, DeRoo LA, et al. Epigenome-wide association study of breast cancer using prospectively collected sister study samples. *J Natl Cancer Inst*. 2013;105(10):694–700.
62. Ulrich CM, Grady WM. Linking epidemiology to epigenomics—where are we today? *Cancer Prev Res (Phila)*. 2010;3(12):1505–1508.
63. Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol Cell*. 2014;54(5):716–727.
64. Nabils NH, Deleyrolle LP, Darst RP, et al. Multiplex mapping of chromatin accessibility and DNA methylation within targeted single molecules identifies epigenetic heterogeneity in neural stem cells and glioblastoma. *Genome Res*. 2014;24(2):329–339.
65. Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet*. 2012;13(5):343–357.
66. Chervona Y, Costa M. Histone modifications and cancer: biomarkers of prognosis? *Am J Cancer Res*. 2012;2(5):589–597.
67. Albert M, Helin K. Histone methyltransferases in cancer. *Semin Cell Dev Biol*. 2010;21(2):209–220.
68. Suvà ML, Riggi N, Bernstein BE. Epigenetic reprogramming in cancer. *Science*. 2013;339(6127):1567–1570.
69. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell*. 2013;152(6):1237–1251.
70. Berdasco M, Esteller M. Genetic syndromes caused by mutations in epigenetic genes. *Hum Genet*. 2013;132(4):359–383.
71. Plass C, Pfister SM, Lindroth AM, et al. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat Rev Genet*. 2013;14(11):765–780.
72. Shen H, Laird PW. Interplay between the cancer genome and epigenome. *Cell*. 2013;153(1):38–55.
73. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190–1195.
74. Kilpinen H, Waszak SM, Gschwind AR, et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science*. 2013;342(6159):744–747.
75. Lee MN, Ye C, Villani AC, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science*. 2014;343(6175):1246980.
76. Ye CJ, Feng T, Kwon HK, et al. Intersection of population variation and autoimmunity genetics in human T cell activation. *Science*. 2014;345(6202):1254665.
77. Farh KK, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337–343.
78. Flanagan JM. Human methylome variation and the rise of epigenetic epidemiology. *Curr Pharmacogenomics Person Med*. 2010;8(2):89–91.
79. Wan ES, Qiu W, Baccarelli A, et al. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Hum Mol Genet*. 2012;21(13):3073–3082.
80. Moghe A, Joshi-Barve S, Ghare S, et al. Histone modifications and alcohol-induced liver disease: are altered nutrients the missing link? *World J Gastroenterol*. 2011;17(20):2465–2472.
81. Bakulski KM, Fallin MD. Epigenetic epidemiology: promises for public health research. *Environ Mol Mutagen*. 2014;55(3):171–183.
82. Heijmans BT, Tobi EW, Stein AD, et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A*. 2008;105(44):17046–17049.
83. Relton CL, Davey Smith G. Is epidemiology ready for epigenetics? *Int J Epidemiol*. 2012;41(1):5–9.
84. Rakyan VK, Down TA, Maslau S, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*. 2010;20(4):434–439.
85. Zakhari S. Alcohol metabolism and epigenetics changes. *Alcohol Res*. 2013;35(1):6–16.
86. Breton CV, Byun HM, Wenten M, et al. Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation. *Am J Respir Crit Care Med*. 2009;180(5):462–467.
87. Breitling LP, Yang R, Korn B, et al. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet*. 2011;88(4):450–457.
88. Javierre BM, Fernandez AF, Richter J, et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res*. 2010;20(2):170–179.
89. Mensaert K, Denil S, Trooskens G, et al. Next-generation technologies and data analytical approaches for epigenomics. *Environ Mol Mutagen*. 2014;55(3):155–170.
90. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet*. 2010;11(3):191–203.
91. Michels KB, Binder AM, Dedeurwaerder S, et al. Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods*. 2013;10(10):949–955.
92. Tsai PC, Spector TD, Bell JT. Using epigenome-wide association scans of DNA methylation in age-related complex human traits. *Epigenomics*. 2012;4(5):511–526.
93. Yu M, Hon GC, Szulwach KE, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*. 2012;149(6):1368–1380.
94. Booth MJ, Branco MR, Ficz G, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*. 2012;336(6083):934–937.
95. Booth MJ, Marsico G, Bachman M, et al. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat Chem*. 2014;6(5):435–440.
96. Ziller MJ, Gu H, Müller F, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 2013;500(7463):477–481.
97. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics*. 2009;10:232.
98. Xi Y, Bock C, Müller F, et al. RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing. *Bioinformatics*. 2012;28(3):430–432.

99. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
100. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011;27(11):1571–1572.
101. Kunde-Ramamoorthy G, Coarfa C, Laritsky E, et al. Comparison and quantitative verification of mapping algorithms for whole-genome bisulfite sequencing. *Nucleic Acids Res.* 2014;42(6):e43.
102. Sun Z, Baheti S, Middha S, et al. SAAP-RRBS: streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing. *Bioinformatics.* 2012;28(16):2180–2181.
103. Sun D, Xi Y, Rodriguez B, et al. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol.* 2014;15(2):R38.
104. Bibikova M, Le J, Barnes B, et al. Genome-wide DNA methylation profiling using Infinium<sup>®</sup> assay. *Epigenomics.* 2009;1(1):177–200.
105. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics.* 2011;98(4):288–295.
106. Paul DS, Beck S. Advances in epigenome-wide association studies for common diseases. *Trends Mol Med.* 2014;20(10):541–543.
107. Wilhelm-Benartzi CS, Koestler DC, Karagas MR, et al. Review of processing and analysis methods for DNA methylation array data. *Br J Cancer.* 2013;109(6):1394–1402.
108. Morris TJ, Beck S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods.* 2015;72:3–8.
109. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008;24(13):1547–1548.
110. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 2014;30(10):1363–1369.
111. Maunakea AK, Nagarajan RP, Bilienky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature.* 2010;466(7303):253–257.
112. Ficiz G, Branco MR, Seisenberger S, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature.* 2011;473(7347):398–402.
113. Riebler A, Menigatti M, Song JZ, et al. BayMeth: improved DNA methylation quantification for affinity capture sequencing data using a flexible Bayesian approach. *Genome Biol.* 2014;15(2):R35.
114. Down TA, Rakyan VK, Turner DJ, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol.* 2008;26(7):779–785.
115. Lienhard M, Grimm C, Morkel M, et al. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics.* 2014;30(2):284–286.
116. Robinson MD, Kahraman A, Law CW, et al. Statistical methods for detecting differentially methylated loci and regions. *Front Genet.* 2014;5:324.
117. Akalin A, Kormaksson M, Li S, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 2012;13(10):R87.
118. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* 2012;13(10):R83.
119. Wang D, Yan L, Hu Q, et al. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics.* 2012;28(5):729–730.
120. Bell JT, Tsai PC, Yang TP, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 2012;8(4):e1002629.
121. Gervin K, Vigeland MD, Mattingsdal M, et al. DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet.* 2012;8(1):e1002454.
122. Liu Y, Aryee MJ, Padyukov L, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol.* 2013;31(2):142–147.
123. Bell JT, Loomis AK, Butcher LM, et al. Differential methylation of the TRPA1 promoter in pain sensitivity. *Nat Commun.* 2014;5:2978.
124. Davies MN, Krause L, Bell JT, et al. Hypermethylation in the ZBTB20 gene is associated with major depressive disorder. *Genome Biol.* 2014;15(4):R56.
125. Yuan W, Xia Y, Bell CG, et al. An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. *Nat Commun.* 2014;5:5719.
126. Aberg KA, McClay JL, Nerella S, et al. MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case-control samples. *Epigenomics.* 2012;4(6):605–621.
127. McClay JL, Aberg KA, Clark SL, et al. A methylome-wide study of aging using massively parallel sequencing of the methyl-CpG-enriched genomic fraction from blood in over 700 subjects. *Hum Mol Genet.* 2014;23(5):1175–1185.
128. van Dijk SJ, Molloy PL, Varinli H, et al. Epigenetics and human obesity. *Int J Obes (Lond).* 2015;39(1):85–97.
129. Rakyan VK, Beyan H, Down TA, et al. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.* 2011;7(9):e1002300.
130. Wong CC, Meaburn EL, Ronald A, et al. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry.* 2014;19(4):495–503.
131. Tobi EW, Slagboom PE, van Dongen J, et al. Prenatal famine and genetic variation are independently and additively associated with DNA methylation at regulatory loci within IGF2/H19. *PLoS One.* 2012;7(5):e37933.
132. Lam LL, Emberly E, Fraser HB, et al. Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A.* 2012;109(suppl 2):17253–17260.
133. Teschendorff AE, Menon U, Gentry-Maharaj A, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* 2010;20(4):440–446.
134. Cheung I, Shulha HP, Jiang Y, et al. Developmental regulation and individual differences of neuronal H3K4me3 epigenomes in the prefrontal cortex. *Proc Natl Acad Sci U S A.* 2010;107(19):8824–8829.
135. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* 2012;13:86.
136. Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics.* 2014;30(10):1431–1439.



137. Zou J, Lippert C, Heckerman D, et al. Epigenome-wide association studies without the need for cell-type composition. *Nat Methods*. 2014;11(3):309–311.
138. Maze I, Shen L, Zhang B, et al. Analytical tools and current challenges in the modern era of neuroepigenomics. *Nat Neurosci*. 2014;17(11):1476–1490.
139. Davies MN, Volta M, Pidsley R, et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol*. 2012;13(6):R43.
140. Revill K, Tycko B. Epigenomic insights into common disease. *Genome Med*. 2011;3(11):71.
141. Gjoneska E, Pfenning AR, Mathys H, et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature*. 2015;518(7539):365–369.
142. Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007;316(5830):1497–1502.
143. Kidder BL, Hu G, Zhao K. ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol*. 2011;12(10):918–922.
144. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
145. Shimbo T, Du Y, Grimm SA, et al. MBD3 localizes at promoters, gene bodies and enhancers of active genes. *PLoS Genet*. 2013;9(12):e1004028.
146. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. *Nat Methods*. 2009;6(11 suppl):S22–S32.
147. Blecher-Gonen R, Barnett-Itzhaki Z, Jaitin D, et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat Protoc*. 2013;8(3):539–554.
148. Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 2012;22(9):1813–1831.
149. Egelhofer TA, Minoda A, Klugman S, et al. An assessment of histone-modification antibody quality. *Nat Struct Mol Biol*. 2011;18(1):91–93.
150. Harvard University. Antibody Validation Database. <http://compbio.med.harvard.edu/antibodies/>. Published 2010. Accessed July 7, 2015.
151. University of California, Santa Cruz. ENCODE Antibodies 2007–2012. <https://genome.ucsc.edu/ENCODE/antibodies.html>. Accessed July 7, 2015.
152. Jung YL, Luquette LJ, Ho JW, et al. Impact of sequencing depth in ChIP-seq experiments. *Nucleic Acids Res*. 2014;42(9):e74.
153. Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*. 2012;13(12):840–852.
154. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, et al. Chromatin state dynamics during blood formation. *Science*. 2014;345(6199):943–949.
155. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
156. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
157. Zang C, Schones DE, Zeng C, et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*. 2009;25(15):1952–1958.
158. Song Q, Smith AD. Identifying dispersed epigenomic domains from ChIP-Seq data. *Bioinformatics*. 2011;27(6):870–871.
159. Marinov GK, Kundaje A, Park PJ, et al. Large-scale quality analysis of published ChIP-seq data. *G3 (Bethesda)*. 2014;4(2):209–223.
160. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*. 2011;147(6):1408–1419.
161. Liang K, Keles S. Detecting differential binding of transcription factors with ChIP-seq. *Bioinformatics*. 2012;28(1):121–122.
162. Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481(7381):389–393.
163. Shen L, Shao NY, Liu X, et al. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One*. 2013;8(6):e65598.
164. Lun AT, Smyth GK. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. *Nucleic Acids Res*. 2014;42(11):e95.
165. Shao Z, Zhang Y, Yuan GC, et al. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol*. 2012;13(3):R16.
166. Bardet AF, He Q, Zeitlinger J, et al. A computational pipeline for comparative ChIP-seq analyses. *Nat Protoc*. 2012;7(1):45–61.
167. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.
168. Bonhoure N, Bounova G, Bernasconi D, et al. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Res*. 2014;24(7):1157–1168.
169. Orlando DA, Chen MW, Brown VE, et al. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep*. 2014;9(3):1163–1170.
170. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, et al. Epigenomic enhancer profiling defines a signature of colon cancer. *Science*. 2012;336(6082):736–739.
171. Paul DS, Soranzo N, Beck S. Functional interpretation of non-coding sequence variation: concepts and challenges. *Bioessays*. 2014;36(2):191–199.
172. Mack SC, Witt H, Piro RM, et al. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature*. 2014;506(7489):445–450.
173. Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet*. 2010;11(7):476–486.
174. Thurman RE, Rynes E, Humbert R, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012;489(7414):75–82.
175. Beyan H, Down TA, Ramagopalan SV, et al. Guthrie card methylomics identifies temporally stable epialleles that are present at birth in humans. *Genome Res*. 2012;22(11):2138–2145.
176. Aberg KA, Xie LY, Nerella S, et al. High quality methylome-wide investigations through next-generation sequencing of DNA from a single archived dry blood spot. *Epigenetics*. 2013;8(5):542–547.
177. Tobi EW, Lumey LH, Talens RP, et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet*. 2009;18(21):4046–4053.

178. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* 2013;41(database issue):D56–D63.
179. Fang F, Turcan S, Rimmer A, et al. Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci Transl Med.* 2011;3(75):75ra25.
180. Schones DE, Cui K, Cuddapah S, et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell.* 2008;132(5):887–898.
181. Henikoff JG, Belsky JA, Krassovsky K, et al. Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci U S A.* 2011;108(45):18318–18323.
182. Valouev A, Johnson SM, Boyd SD, et al. Determinants of nucleosome organization in primary human cells. *Nature.* 2011;474(7352):516–520.
183. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc.* 2010;2010(2):pdb.prot5384.
184. He HH, Meyer CA, Hu SS, et al. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat Methods.* 2014;11(1):73–78.
185. Giresi PG, Kim J, McDaniell RM, et al. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 2007;17(6):877–885.
186. Song L, Zhang Z, Grasfeder LL, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 2011;21(10):1757–1767.
187. Simon JM, Hacker KE, Singh D, et al. Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. *Genome Res.* 2014;24(2):241–250.
188. Buenrostro JD, Giresi PG, Zaba LC, et al. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10(12):1213–1218.
189. Buenrostro JD, Wu B, Chang HY, et al. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol.* 2015;109:21.29.21–21.29.29.
190. Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell.* 2007;129(4):823–837.
191. Leung D, Jung I, Rajagopal N, et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature.* 2015;518(7539):350–354.
192. Meissner A, Gnirke A, Bell GW, et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 2005;33(18):5868–5877.
193. Bock C, Tomazou EM, Brinkman AB, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol.* 2010;28(10):1106–1114.
194. Jacinto FV, Ballestar E, Esteller M. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *Biotechniques.* 2008;44(1):35, 37, 39 passim.

---

## APPENDIX

### Suggested Resources

National Institutes of Health Common Fund Epigenomics Program:

- <http://commonfund.nih.gov/epigenomics/grants>

International Human Epigenome Consortium (the goal is to generate 1,000 reference epigenomes):

- <http://ihec-epigenomes.org/>

Encyclopedia of DNA Elements (ENCODE) Consortium (extensive epigenome data from cultured cell lines):

- <http://www.genome.gov/ENCODE/>

National Institutes of Health Roadmap Epigenomics Mapping Consortium (epigenomic maps for stem cells and primary ex vivo tissues):

- <http://www.roadmapepigenomics.org/>
- [http://egg2.wustl.edu/roadmap/web\\_portal/](http://egg2.wustl.edu/roadmap/web_portal/)

European BLUEPRINT project (the goal is to generate approximately 100 reference epigenomes):

- <http://www.blueprint-epigenome.eu/>

International Cancer Genome Consortium (include epigenomic maps in 50 different tumor types or subtypes):

- <https://icgc.org/>

The Cancer Genome Atlas:

- <https://tcga-data.nci.nih.gov/tcga/>

Epigenomics mirror of the University of California, Santa Cruz, genome browser:

- <http://www.epigenomebrowser.org/>

National Center for Biotechnology Information epigenomics portal:

- <http://www.ncbi.nlm.nih.gov/epigenomics>

Galaxy (for data-intensive biomedical research):

- <http://galaxyproject.org/>