Chapter 10
Clinical Epigenetics and Epigenomics

Chuan Qiu, Fangtang Yu, Hong-Wen Deng, and Hui Shen

Abstract Epigenetics is the study of somatically heritable changes in gene expression that occur without alterations in DNA sequence, mainly including DNA methylation and histone modification. Epigenomics refers to the complete study of these somatically heritable changes across the whole genome. Epigenetic mechanisms are critical components in the growth of cells and normal development. Aberrant epigenetic changes have been found to be causative factors in cancer, autoimmune diseases as well as others. Significant progress has been made towards epigenomic profiling by using molecular techniques. In this chapter, we introduce the basics of epigenetics and epigenomics; describe the remarkable advances in current epigenomic mapping and analysis technologies, especially microarray-based and next-generation sequencing-based applications. Then we focus on the recent studies of epigenetic changes in normal and diseased cells with the aim to translate basic epigenetic and epigenomics research into clinical applications. We also discuss some critical challenges ahead and provide a perspective on the progress of epigenomics field.

Keywords Epigenetics • Epigenomics • DNA methylation • Histone modification • Mapping and analyzing technologies • Human disease

Abbreviations

5-hmC	5-hydroxymethyl-cytosine
5-fC	5-formyl-cytosine
5-mC	5-methylcytosine
5-caC	5-carboxylcytosine

C. Qiu • F. Yu • H. Shen (✉)
Center for Bioinformatics & Genomics, Department of Biostatistics & Bioinformatics, School of Public Health & Tropical Medicine, Tulane University, New Orleans, LA 70112, USA
e-mail: hshen3@tulane.edu

H.-W. Deng
Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA
e-mail: hdeng2@tulane.edu

CGI   CpG island
DMH   differential methylation hybridization
DMRs   differentially methylated regions
DNMTs   DNA methyltransferases
EWAS   epigenome-wide association studies
FDR   false discovery rate
HATs   histone acetyltransferases
HDACs   histone deacetylases
HMTs   histone methyltransferases
IMA   Illumina Methylation Analyzer
LOESS   locally weighted scatterplot smoothing
MBD   methyl-CpG binding domain
MeCPs   Methyl-CpG Binding Proteins
MeDIP   methylated DNA immunoprecipitation
PRC   polycomb repressive complexes
RRBS   reduced representation bisulfite sequencing
SNPs   Single Nucleotide Polymorphisms
TET   ten-eleven translocation
UHRF   ubiquitin plant homeodomain RING finger
WGBS   whole genome bisulfite sequencing

In 1959, Waddington first coined the concept "epigenetics" (Waddington 1959), which now refers to the mechanism for stable maintenance of gene expression changes that involves physically "marking" DNA or its associated proteins other than alterations in DNA sequence.

A variety of epigenetic factors have been identified, such as DNA methylation, histone modification, and non-coding RNAs (e.g., microRNAs and long non-coding RNAs) etc. These epigenetic factors coordinatively regulate gene expression and provide heritable epigenetic information that is not encoded in DNA sequence (Cedar and Bergman 2009; Esteller 2011). Epigenome refers to the entire constitution of epigenetic marks in a cell type at a given time point, it is cell-specific and tissue-specific (Varley et al. 2013). In each type of cells, epigenetic factors regulate gene expression in different ways, for example, facilitate or restrict transcription factor access to DNA sequence (Rivera and Ren 2013). Epigenome may change over the lifetime (Fraga et al. 2005a) and are prone to environmental influences, such as stress, social interactions, physical activity, exposure to toxins and diet (Alegria-Torres et al. 2011). Aberrant epigenomic alternations have been implicated in a wide variety of human disorders, such as cancer and autoimmune diseases etc. (Portela and Esteller 2010), and epigenetic drugs may revolutionize the treatment of many human diseases (Heerboth et al. 2014). In this chapter, we briefly reviewed the molecular basis for two major epigenetic factors, DNA methylation and histone modification, and discussed some commonly used epigenome-wide analytic approaches for these two factors, as well as their involvement in some human complex disorders.

10.1 Molecular Basis of DNA Methylation and Histone Modification

10.1.1 DNA Methylation

DNA methylation, commonly called the 'fifth base' in the genome, is one of the most extensively studied epigenetic mechanisms. It is a direct chemical modification of the fifth carbon of a cytosine that adds a methyl ($-CH_3$) group through a covalent bond resulting in 5-methylcytosine (5-mC).

In adult somatic tissues, DNA methylation typically occurs in a "CpG" (C-phosphate-G) dinucleotide context (Bird and Southern 1978; Cedar et al. 1979). An exception to this is seen in embryonic stem cells (Haines et al. 2001), where a substantial amount of 5-mC is also observed in non-CpG sites (mCHG, mCHH). In human genome, there are 28 million CpG sites, which are not evenly distributed throughout the genome (Lister et al. 2009) but tend to cluster in regions, known as "CpG islands" (CGIs). CGIs usually occur near gene transcription start sites (TSS) and ~60 % of human gene promoters are associated with CGIs (Bird 1986; Gardiner-Garden and Frommer 1987). DNA methylation is catalyzed by a family of enzymes termed DNA methyltransferases (DNMTs) (Okano et al. 1998), including DNMT1, DNMT3A, DNMT3B, DNMT2 and DNMT3L, which cooperate in establishing and maintaining DNA methylation patterns (Kulis and Esteller 2010; Okano et al. 1999). Equally important and opposite with DNA methylation is DNA demethylation. DNA demethylation can be either passive or active, or a combination of both. Passive DNA demethylation usually takes place when DNMT1 cannot effectively restore the DNA methylation patterns on newly synthesized DNA strands during replication rounds (Wu and Zhang 2010), whereas active demethylation is usually mediated by the ten-eleven translocation (TET) family enzymes (TET1, TET2 and TET3) and subsequent restoration of unmodified cytosine by the thymine DNA glycosylase (TDG)-mediated base excision repair (Kohli and Zhang 2013).

The importance of DNA methylation as a major epigenetic modification in gene expression has been widely recognized. Hypermethylation of CpGs in TSS proximal regions, particularly in promoter CGIs, is largely associated with repressed gene transcription (Wagner et al. 2014), whereas methylation of CpGs located within gene bodies is usually associated with an increase in transcriptional activity (Ramsahoye et al. 2000; Hellman and Chess 2007). However, several recent studies have revealed that there is no simple relationship between inter-individual DNA methylation and gene expression with respect to the location of the methylated CpGs and both negative and positive inter-individual methylation-expression correlations were detected for CpGs located in gene body and transcription start site proximal regions, as well as in intergenic regions (Wagner et al. 2014; Bell et al. 2011).

10.1.2 Histone Modification

The basic unit of chromatin is the nucleosome, which is composed of an octomer of histone proteins (containing two copies each of histones H2A, H2B, H3, and H4) around which is wrapped a length of 147 bp DNA. The degree to which chromatin are condensed or packed is a critical determinant of the transcriptional activity of the associated DNA and this is mediated in part by diverse post-translational covalent modifications of the N-terminal tails of histone proteins (Fig. 10.1).
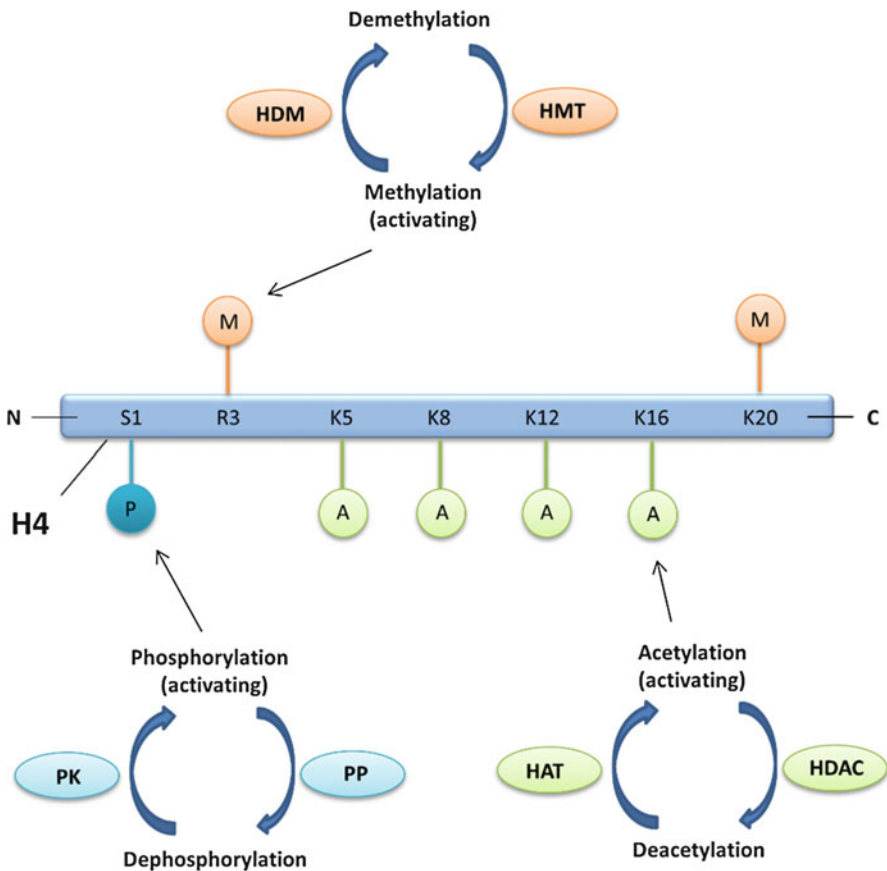

Fig. 10.1 Post-translational modifications of histones. The first 20 amino acids in the N-terminus of the human histone H4 are illustrated. Many sites in the N-terminus can be targets for epigenetic tagging such as acetylation (A), phosphorylation (P) and methylation (M). Acetylation is catalyzed by histone acetyltransferase (HAT) and removed by histone deacetylase (HDAC); Phosphorylation is catalyzed by protein kinases (PK) and removed by protein phosphatase (PP); Methylation is catalyzed by histone methyltransferases (HMT) and removed by histone demethylase (HDM). Some histone modification marks are associated with gene activation while others are associated with gene repression, and the integration of multiple marks leads to a finely tuned transcriptional response

At least eight different types of histone modification have been identified: acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP ribosylation, deimination, and proline isomerization. All the modifications are reversible and dynamic, mediated by enzymes that add/remove modification.

Histone acetylation occurs via an enzymatic transferring of an acetyl group from acetyl-CoA to the ε-NH+ group of the lysine residues within a histone. This enzymatic activity is catalyzed by enzymes called histone acetyltransferases (HATs) and reversed by histone deacetylases (HDACs) (Hodawadekar and Marmorstein 2007). Histone acetylation is a hallmark of transcriptional activation (Sterner and Berger 2000) and the histone acetylation patterns are tightly associated with many cellular processes including chromatin dynamics and transcription, gene silencing, cell cycle progression, apoptosis, differentiation, DNA replication, DNA repair, nuclear import, and neuronal repression (Cohen et al. 2011).

Histone methylation is another extensively studies histone modification marks. It is defined as the transfer of one, two, or three methyl groups from S-adenosyl-L-methionine to lysine or arginine residues of histone proteins by histone methyltransferases (HMTs). In the cell nucleus, when histone methylation occurs, specific genes within the DNA complexed with the histone may be activated or silenced (Greer and Shi 2012). For instance, the tri-methylation of histone H3 at lysine 4 (H3K4me3) is positively correlated with gene transcription and commonly detected in a tight, localized area at 5′-ends/promoter regions of active genes (Barski et al. 2007). H3K36me3 is strongly enriched across the gene body and at the 3′-end of active genes and may link to transcriptional elongation (Barski et al. 2007). In contrast, H3K27me3 is the classic repressive histone modification mark, which shows a broad peak at promoters and throughout the gene body of the silent genes (Barski et al. 2007).

10.2 Epigenome-Wide Analyses of DNA Methylation and Histone Modification

10.2.1 Epigenome-Wide DNA Methylation Analysis

10.2.1.1 DNA Methylation Profiling Assays

DNA methylation analysis normally relies on three strategies (Fig. 10.2): (1) Digestion of genomic DNA with methylation-sensitive restriction enzymes; (2) Affinity-based enrichment of methylated DNA fragments; and (3) Bisulfite conversion. Each of the three strategies can be combined with either microarray or next-generation sequencing technique to interrogate epigenome-wide DNA methylation patterns, and each with unique advantages and drawbacks (Table 10.1).

- **Digestion of genomic DNA with methylation-sensitive restriction enzymes**: Some restriction enzymes (e.g., HpaII and SmaI) are methylation sensitive –
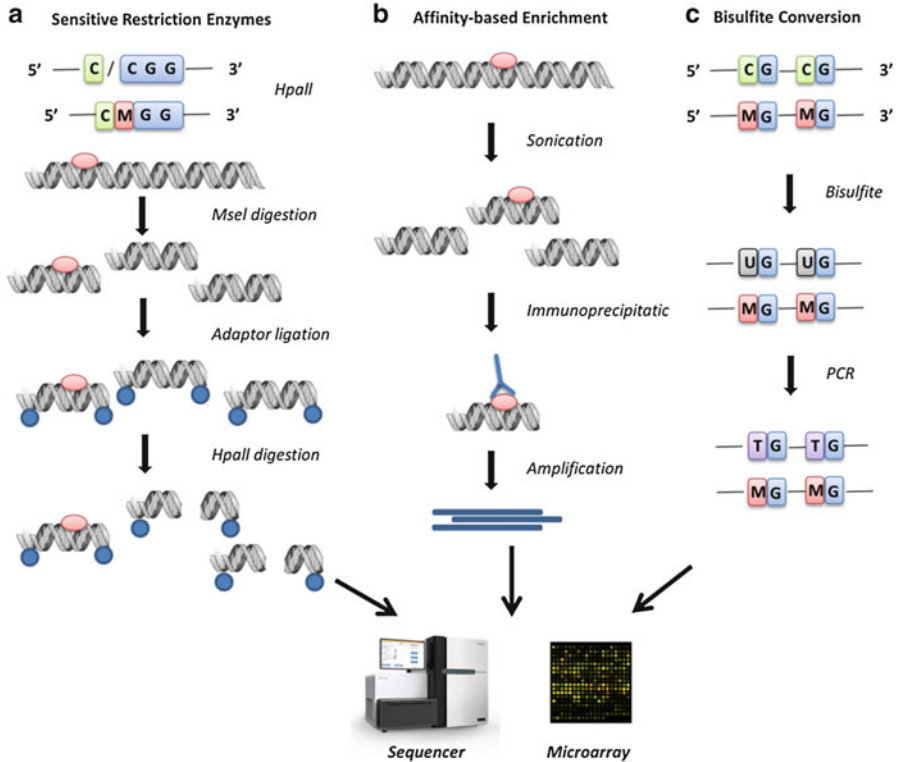
Fig. 10.2 Strategies for pretreatment of DNA sample. (**a**) Digestion of genomic DNA with methylation-sensitive restriction enzymes. (**b**) Affinity-based enrichment of methylated DNA fragments. (**c**) Chemical treatment of DNA with sodium bisulfite results in the conversion of unmethylated cytosines to uracils. In contrast, methylated cytosines are protected. Subsequently, microarray or next-generation sequencing of these libraries reveals the methylation status

their activity is affected by the presence of a methyl CpG within restriction sites. Therefore, when genomic DNA is digested with a methylation-sensitive restriction enzyme, difference in methylation status is converted into difference in sequence fragment size. For example, differential methylation hybridization (DMH) uses combinations of methylation-sensitive and methylation-insensitive restriction enzyme digestion, followed by ligation-mediated PCR to enrich for methylated or unmethylated fragments. PCR products are labeled and hybridized to arrays, or tested by next-generation sequencing (Fig. 10.2a). Methods based on this strategy were used in early epigenome-wide DNA methylation studies (Rakyan et al. 2011), but the genome-wide CpG coverage and resolution are limited by the cutting frequency and the fragment size of the chosen restriction enzymes.

- **Affinity-based enrichment of methylated DNA fragments**: Affinity-based enrichment assays capture methylated DNA fragments with a methyl-CpG

Table 10.1 Comparison of DNA methylation detection procedure

	Advantages	Disadvantages	Suitable application	Methods
Affinity-based enrichment	Rapid and efficient genome-wide assessment of DNA methylation	Possibility of antibody cross-reactivity	Rapid, large scale, low resolution study of DNA methylation	MeDIP-seq
	Powerful tool for comprehensive profiling of DNA methylation in complex genomes	Resolution depends on the fragment size of the enriched methylated DNA		MBD-seq
		Does not provide single base pair resolution		
Bisulfite conversion	Resolution at the nucleotide level	Often leads to damaged DNA	High resolution study of DNA methylation at small or large scale	WGBS
	Effectively converts an epigenetic difference into a genetic difference, easily detectable by sequencing	Potentially incomplete conversion of DNA		RRBS
		Cannot distinguish 5-mc and 5-hmc		Infinium
		Whole genome sequencing requires intensive downstream analysis		
Restriction enzyme-based digestion	Easy to use	Determination of methylation status is limited by the enzyme recognition site	Targeted, site specific study of DNA methylation	DMH
	High enzyme turnover			MSDK

Note: The table is modified from Ku et al. (2011)

binding protein (e.g., MBD2) or 5-mC specific antibody (Fig. 10.2b). For example, when performing methylated DNA immunoprecipitation (MeDIP)-chip/-seq, genomic DNAs are first fragmented by sonication and then incubated with anti-5mC antibody. The anti-5mC bound DNA fragments are isolated, deproteinized, and then hybridized onto microarrays (MeDIP-chip) or analyzed by next-generation sequencing (MeDIP-seq). The frequency of DNA fragments bound to specific probes or mapped to specific genomic regions provides the raw data from which DNA methylation levels can be inferred. The affinity-based DNA methylation assays allow for rapid and efficient genome-wide assessment of DNA methylation, however, as the affinity-captured DNA fragments are generally hundreds of nucleotides in size, the major limitation of these methods is its inability to pinpoint methylation changes at a single CpG resolution (Robinson et al. 2010).

- **Bisulfite Conversion**: Bisulfite conversion of DNA is the most commonly used method for DNA methylation studies. It uses bisulfite salt to deaminate cytosine residues on single-stranded DNA, converting them to uracil while leaving 5-methylcytosine intact. Once a difference of methylation status is converted into a difference of DNA sequence, it can be detected by various techniques (Fig. 10.2c). Bisulfite sequencing applies routine DNA sequencing methods on bisulfite-converted genomic DNA. It can provide quantitative methylation measurement at single nucleotide resolution and is widely accepted as a gold standard for DNA methylation analysis. Recent development of next-generation sequencing technology makes it feasible to perform whole genome bisulfite sequencing (WGBS) (Suzuki and Bird 2008). Though WGBS can provide a comprehensive coverage of almost all CpGs in the human genome, its usage is currently limited by its high cost. Thus, several more cost-effective bisulfite conversion-based approaches, such as reduced representation bisulfite sequencing (RRBS) and Illumina 450 k array, are widely employed in the current epigenomics research field. In RRBS, genomic DNA is digested by the methylation-insensitive restriction enzyme MspI (5′-C′CGG-3′) and separated by gel electrophoresis, and then size-selected DNA fragments are bisulfite converted and analyzed by next-generation sequencing platforms (Meissner et al. 2005). After the MspI digestion and size selection, CpG sites were enriched in sequencing library that reduce the amount of nucleotides needed to be sequenced.

10.2.1.2   Data Processing and Analysis

In this section, we will review the data processing approaches for three most popular epigenome-wide DNA methylation profiling methods, namely, Illumina 450k array, RRBS and MeDIP-seq. We also discussed the approaches for data visualization and identification of differential methylation in the following section.

Data Processing for Illumina 450k Array

The Illumina 450k array adapted BeadArray technology to recognize the bisulfite-converted DNA for interrogation of DNA methylation. It offers a unique combination of comprehensive, expert-selected coverage, high sample throughput and an affordable price, making it the most widely used method for current epigenome-wide association studies (EWAS). The Illumina 450k array tests more than 485,000 CpGs at single-nucleotide resolution, which covers 99 % of RefSeq genes and 96 % of CGIs (Bibikova et al. 2011). The data processing procedures for Illumina 450k array contain several main steps, including quality control (QC), normalization, adjustment of batch effect, and calculation of DNA methylation levels.

– **QC**: The aim of this step is to detect and filter out samples and probes that do not meet the experimental standard. The Illumina 450k arrays contain several control probes for determining the data quality. Diagnostic plots of control probes in Illumina Genome-Studio program can be used to detect poorly performed samples (Bibikova et al. 2011). Assessing for poor quality samples can also be carried out by functions embedded in several R-packages specifically for analyzing Illumina 450k arrays, such as HumMethQCReport (Mancuso et al. 2011), IMA (Wang et al. 2012), Minfi (Aryee et al. 2014) and MethyLumi (Davis et al. 2012).

For QC of probes, some packages such as IMA (Wang et al. 2012) filter out probes for which a large proportion of samples (i.e., >25 %) have a detection P-value >0.05. LumiWCluster avoids to discard probes (Kuan et al. 2010), instead it incorporates all the data while accounting for the quality of individual observations. A particular issue for QC of 450k array is that certain probes contain single nucleotide polymorphisms (SNPs) within the targeted sequences and thus the methylation levels assessed by these probes may be influenced by the DNA genotype (Dedeurwaerder et al. 2011). Hence, several programs (e.g., IMA) have incorporated functions to filter out these SNP-associated CpG probes (Wang et al. 2012; Touleimat and Tost 2012).

– **Normalization**: Normalization step is used to remove technical and systematic variation which could mask true biological differences. There are two types of normalization approach: (1) between-array normalization: address the comparability of intensity distribution between multiple arrays; (2) within-array normalization: correction for dye, intensity and spatial dependent bias within individual arrays (Siegmund 2011). The Illumina GenomeStudio uses a basic normalization approach by treating the first sample in the array as the reference but allows the user to reselect the reference sample if the first sample shows poor quality. This approach is also implemented in R-package MethyLumi (Davis et al. 2012) and Minfi (Aryee et al. 2014). Locally weighted scatterplot smoothing (LOESS) and quantile normalization assume similar total methylation signals across samples and may potentially discard the true biological signals (Laird 2010). There also exist several other approaches for normalizing the probe intensities (Marabita et al. 2013), but currently a lack of consensus exists regarding to the optimal normalization algorithm.

• **Adjustment of Batch Effect**: Batch effects represent measurements that have different behavior across conditions but are not related to the biological or scientific questions in a study (i.e. experiment time, chip or instrument used and laboratory conditions.). Some of the factors can be corrected by careful study design, for example, equally splitting the cases and controls into different batches by random sampling (Johnson et al. 2007) Other potential confounders may be corrected by several computational methods. For example, R-packages ComBat is a widely used adjustment method. It is based on empirical Bayes procedure (Johnson et al. 2007) and is robust to outliers in small sample sizes (Sun et al. 2011).

– **Calculation of DNA Methylation Levels**: DNA methylation levels are determined based on the intensities of the fluorescence signals from probes. The main output is the β-value and M-value which are ready for downstream statistical analysis. The β-value is calculated with the intensity of signal from methylated alleles (Max(M,0)) and the intensity of signal from unmethylated alleles (Max(U,0)) by the following formula:

$$\beta = \frac{\text{Max}(M, 0)}{\text{Max}(M, 0) + \text{Max}(U, 0) + 100}$$

The obtained β-value denotes the average methylation level for each CpG site. It ranges from 0 (unmethylated) to 1 (fully methylated) on a continuous scale. Alternative, some researchers use M-value to indicate the methylation level, which is calculated as

$$M = \log_2 \frac{\text{Max}\,(M, 0) + 1}{\text{Max}(U, 0) + 1}$$

The range of M-values is negative infinity to positive infinity, which is consistent with data from normal distribution. However, the interpretation is of M-values is not as intuitive as β-value. The relationship of M-values and β-value is:

$$M = \log_2 \frac{\beta}{1 - \beta}$$

Thus, positive M-values correspond to a methylation rate greater than 50 %, while negative M-values indicate a methylation rate less than 50 %.

Data Processing for RRBS

Processing of RRBS data mainly involves two steps, QC and alignment of sequencing reads.

– **QC**: The raw sequencing reads are normally generated in the fastq format, which records the sequence of nucleotides and their base call confidence levels. In order to obtain high quality RRBS data, several technical details require careful attention. For example, the incomplete bisulfite conversion will lead to spuriously elevated DNA methylation levels. One should use spike-in control DNAs with known DNA methylation levels to monitor the sensitivity and specificity of bisulfite conversion. Alternatively, elevated levels of observed CpC methylation can also provide an indication of incomplete bisulfite conversion because CpC dinucleotides are rarely methylated in mammalian cells (Bock 2012). Some of the QC steps for the RRBS data can be performed by QC tools (e.g., NGS QC

toolkit) (Patel and Jain 2012) that are generally applicable to the next-generation sequencing produced reads, while other QC criteria such as efficiency of bisulfite conversion require QC tools that are dedicated to bisulfite sequencing, such as BSeQC (Lin et al. 2013).

– **Alignment**: Because of the reduced sequence complexity of the bisulfite converted sequence reads, alignment of bisulfite converted sequence reads to the reference genome require specific alignment tools. Generally, the alignment tools can be categorized into two groups: three-letter aligners and wild-card aligners. Bismark (Krueger and Andrews 2011) and BS-Seeker (Chen et al. 2010) are examples of three-letter aligners, which convert C to T in both sequenced reads and reference sequences prior to alignment. In contrast, wild-card aligners like BSMAP/RRBSMAP (Xi and Li 2009; Xi et al. 2012) replace Cs in the sequenced reads with wild-card Y but do not need the reference genome conversion step. Compare with whole-genome bisulfite alignment tool, such as an extensively validated MAQ-based pipeline, these specific aligners (e.g. RRBSMAP) could maintain high mapping accuracy and consistency between replicates, and also significantly improve runtime performance and memory efficiency (Xi et al. 2012).

– **Calculation of DNA Methylation Signals**: As unmethylated cytosines will be converted to Ts by the bisulfite treatment and methylated cytosines will stay Cs, absolute DNA methylation level could be calculate by counting the number of Cs and Ts at each C and simply divide the number of Cs by the total number of Cs and Ts.

Data Processing for MeDIP-Seq

In MeDIP-seq, the information of enrichment or depletion of extended sequencing reads will be used to estimate the methylation level of specific regions in the genome, the reads sequence itself does not provide methylation information. As a result, specific data processing approaches are needed to estimate the DNA methylation levels from MeDIP-seq method.

– **QC and alignment**: similar to other sequencing-based methods, the first step in the analysis of MeDIP-/MBD-seq is QC and alignment of sequencing reads to the reference genome, which can be conducted by using a standard quality control program and aligner, such as Bowtie2 (http://bowtie-bio.sourceforge. net/bowtie2/index.shtml) and BWA (Li and Durbin 2009).

– **Estimation of DNA methylation levels**: after alignment, the unique mapped reads are then extended to MeDIP-enriched DNA fragment size, the DNA sequence of each chromosome is divided into a series of certain base pair intervals (e.g. 50bp), and the extended reads in each interval are counted as the methylation signal in this region. These estimated DNA methylation signal can be confounded by varying density of methylated CpG sites. That is, regions with high CpG densities can give rise to high enrichment scores even with low absolute DNA

methylation levels and low CpG density regions can produce low enrichment scores even with high levels of DNA methylation. Down et al. developed the tool BATMAN which applies a Bayesian method to estimate absolute methylation values from MeDIP-chip or MeDIP-seq data (Down et al. 2008). It provides accurate estimations of methylation value, however it is not especially user-friendly and is quite a computationally technical process. Another tool is R package MEDIPS, it is a comprehensive approach for normalizing and analyzing MeDIP-seq data (Chavez et al. 2010). This method is based on the valuable concept of coupling factors presented by BATMAN (Down et al. 2008). MEDIPS incorporates a statistical frame work developed for count data which models the read number by an overdispersed Poisson model. This method could significantly reduce run time for processing MeDIP-seq data and easy to use.

10.2.1.3 Identifying Differentially Methylated Regions (DMRs)

In clinical study (e.g. case and control study), it is crucial important to identify the DMRs between different experimental condition. There are several different types of DMR, such as tissue-specific DMR and aging-specific DMR (Rakyan et al. 2011). According to DNA methylation profiling methods we use, these DMR can be a single CpG site or a region of interest (e.g. promoters, CGIs). The Student's t-test and Wilcoxon rank sum test can be used to identify DMRs by using the normalized methylation signal between two groups. Bock, C (Bock 2012) well summarized several other advanced methods which aim to improve DMR detection (e.g. mixture models (Wang 2011), stratification of t-test (Chen et al. 2012) and point out it is difficult to predict which methods will work best for real-world DNA methylation data sets. There are several different tools used for identification of DMRs. For Illumina 450k array, most commonly used tools including R package IMA (Wang et al. 2012) and Minfi (Aryee et al. 2014) etc. The IMA (Wang et al. 2012) apply Student's t-test and empirical Bayes statistics, it allows identification of DMRs in both single CpG sites and regions of interest. For regions of interest differential methylation analysis, IMA will compute the mean, median or Tukey's biweight robust average for the loci within that region and create an index. limma uses an empirical Bayes moderated t-test to improve power in small sample sizes. M-values should be used in these cases as they will rely much more heavily on the assumption of normality. Minfi (Aryee et al. 2014) uses an F-test or linear regression to test each genomic position for association between methylation and categorical or continuous phenotype, respectively. R package methylKit (Akalin et al. 2012) is most commonly used tools for RRBS data analysis. It applies a t-test or logistic regression to calculate p-values which are adjusted to q-values for multiple test correction. For MeDIP-seq data, R package MEDIPS is sufficiently fast and could be practical for routine processing of MeDIP–seq (Bock 2012). Importantly, we need to concern the issue of correction multiple hypothesis testing since the tests for differential DNA methylation are performed simultaneously at a large number of genomic loci.

10.2.2   Epigenome-Wide Histone Modification Analysis

10.2.2.1   Histone Modification Profiling Assays

Methods for epigenome-wide analysis of histone modification marks rely heavily on a procedure called chromatin immunoprecipitation (ChIP). The basic steps of ChIP includes: (1) Crosslink DNA and associated proteins on chromatin in cells; (2) Sonicate the DNA-protein complexes into ~500 bp fragments; (3) Immuno-precipitate DNA fragments using specific antibody against the particular histone mark; (4) Purify the immunoprecipitated DNA fragments and subsequently analyze by microarrays (ChIP-chip) or sequencing (ChIP-seq) (Fig. 10.3). To control for the effects of non-specific bindings, nonspecific immunoglobulin G (IgG) antibodies and input chromatin have been commonly used as controls (Kidder et al. 2011). Regions showing enrichment of ChIP products over controls represent DNA sequences where the specific histone modification marks are associated with in vivo. In addition to histone modification marks, the ChIP-chip/-seq methods can also be used to map global binding sites for specific transcription factors, RNA polymerases, or in principle any DNA-associated proteins.

10.2.2.2   Data Processing and Analysis

Using standard QC and alignment programs, the high quality sequencing reads from ChIP-seq data can be selected and mapped to the reference genome. The aligned reads are then used to identify regions of increased read tag density relative to the background estimated from the IgG/input controls. One straightforward approach is simply to use a minimum fold enrichment threshold of ChIP tags over normalized control tags in candidate regions/tiling windows. However, any threshold is arbitrary and prone to error, this approach does little to assist the user in assessing the significance of peaks (Wilbanks and Facciotti 2010). More sophisti-cated statistical approaches have been incorporated to identify and assess the significance of putative peaks (Pepke et al. 2009). So far, over 40 different 'peak calling' programs have been developed under a variety of statistical models, such as Poisson, local Poisson, t-distribution, conditional binomial, and hidden Markov models. Though a few studies attempted to compare the performance of some of these peak calling programs (Wilbanks and Facciotti 2010; Micsinai et al. 2012), there does not appear to be a clear winner and many program have multiple parameters that can be adjusted by the user. As using different programs or different parameter settings can significantly affect the final peak lists, care must be taken that data sets that are to be compared must be analyzed using the same methods and settings.
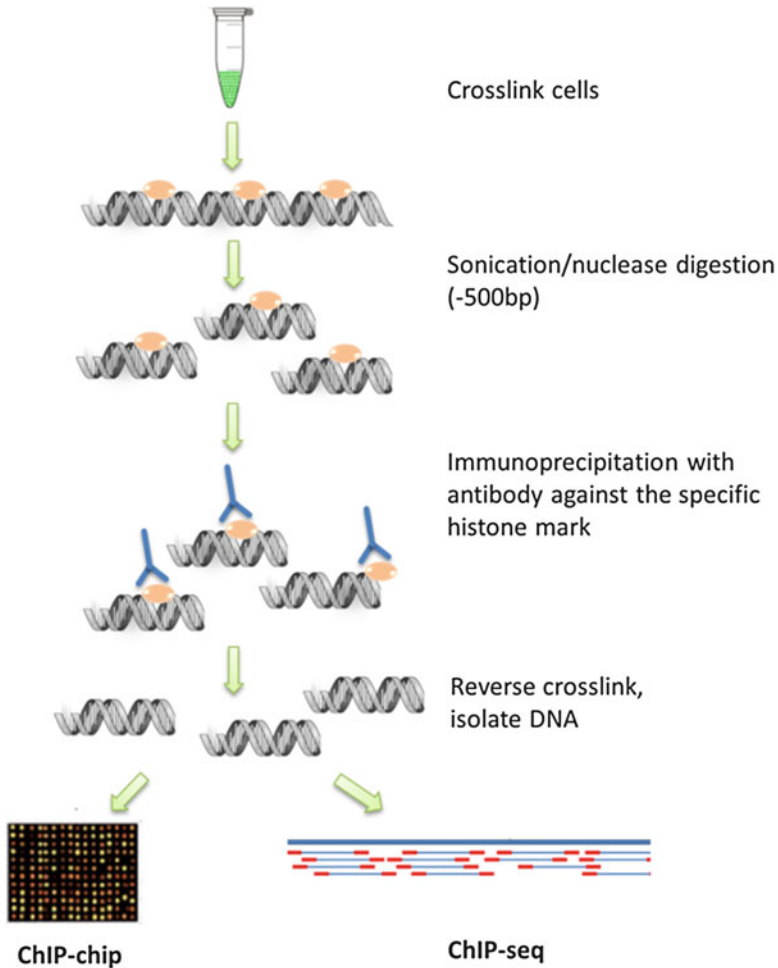
Fig. 10.3 Workflow for ChIP-chip and ChIP-seq. DNA and associated proteins are crosslinked and sheared into ~500 bp DNA fragments by sonication or nuclease digestion, DNA fragments associated with the histone mark of interest are selectively immunoprecipitated using an antibody specifically against the particular histone mark. Purified DNA can be analyzed by microarrays (ChIP-chip) or sequencing (ChIP-seq)

10.2.3   Epigenomic Data Visualization and Interpretation

Visualization of DNA methylation data (e.g. MeDIP-seq, RRBS) and histone modification (ChIP-seq) data is incredibly important. It enables you to investigate the data and may help you come up with new ideas about how to analyze the data. The ability to visualize these kinds of data is enabled through the use of some popular genome browsers, such as UCSC Genome Browser (Kent et al. 2002) and Integrative Genome Viewer (IGV) (Robinson et al. 2011). UCSC Genome Browser

includes lots of published studies and ENCODE data, it is useful for data integration and visualization. However, the data files need to be uploaded to this web-based genome browser which makes a little more difficult to upload large custom data sets. IGV is Java based genome browser. It runs locally on your own computer. It does not have the same degree of public available data as UCSC genome browser, but tend to be somewhat faster for browsing across the genome. Also, it is better for looking at individual reads. There are several types of file format, such as BED, Wiggle and bedGraph format. BED files are very basic as they simply describe a region in the genome. They are usually used to describe MeDIP-seq and ChIP-Seq peaks. Nearly every genome browser supports visualization of BED files. Wiggle files are used to display quantitative information across genomic regions. Wiggle format is compact and displays data at regular intervals. Similar to Wiggle format, bedGraph use variable length intervals instead of constant intervals found in wiggle files, and are usually a little bigger in size. There are a bunch of specialized programs for creating genome browser files, such as bedToBigBed (https://www.encodeproject.org/software/bedToBigBed/) and igvtools (https://www.broadinstitute.org/igv/igvtools).

Several bioinformatics tools were used to interpret biological meaning from epigenomic data results. For example, EpiExplorer (Halachev et al. 2012) empowers biologists to explore large epigenome datasets in real time and over the Internet. It facilitates interactive hypothesis generation and identification of candidates for experimental follow-up. Cytoscape (Shannon et al. 2003) is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. DAVID (da Huang et al. 2009) provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes.

10.3   Epigenomics of Human Diseases

With the advent of new technologies we are starting to unravel the epigenomic mechanisms underlying a diverse range of human disorders, such as cancer and autoimmune diseases. A comprehensive understanding of epigenetic mechanisms, their interactions and alterations in human disease, has become a priority in clinical research (Portela and Esteller 2010).

10.3.1   Epigenomics of Cancer

Diverse altered DNA methylation patterns have been implicated in the pathogenesis and metastasis of various cancers. Genome-wide hypomethylation has been revealed in several common cancer types, such as stomach, liver and lung cancers

(Kulis and Esteller 2010) DNA hypomethylation occurs mostly at DNA-repetitive regions resulting in activation of genes with growth and tumor promoting functions and loss of genome stability and imprinting (Esteller 2008). A clear case is the long interspersed nuclear element (LINE) family member LINE1. Many studies have support correlations between LINE1 hypomethylation and increased risk of cancer (Barchitta et al. 2014). For example, hypomethylation of a specific LINE-1 promoter was found to induce an alternate transcript of the MET oncogene in bladder tumors and across the entire urothelium of tumor-bearing bladders (Wolff et al. 2010). A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer (Antelo et al. 2012; Ogino et al. 2013), and hypomethylation of LINE-1 in primary tumor has been associated with poor prognosis and survival in young breast cancer patients (van Hoesel et al. 2012) and prominent hypomethylation of Alu and LINE-1 in HER2 enriched subtype may be related to chromosomal instability (Park et al. 2014). In addition to the effects on repetitive elements, promoter hypomethylation can activate the aberrant expression of oncogenes and result in loss of imprinting in some loci (Portela and Esteller 2010). For instance, loss of imprinting of IGF2 gene has been associated with an increased risk of different types of cancer (Lim and Maher 2010). Recent study also shows that hypomethylation in TP73 and TERT gene body alter the transcriptional landscape of growth rate of glioblastoma through the activation of a limited number of normally silenced promoters within gene bodies, result in activating the aberrant expression of an oncogenic protein (Nagarajan et al. 2014). Hypermethylation at the CGIs of certain promoters causing transcriptional silencing of tumor suppressor gene were also observed. The transcriptional silencing caused by promoter hypermethylation affects genes involved in the multiple cellular pathways (Portela and Esteller 2010), such as DNA repair (e.g., MGMT, MLH1, MSH2, GSTP1), Ras signaling (e.g., DAPK, NOREIA, RASSFIA, RECK) etc. (Esteller 2007). For example, hypermethylation at CGI of MLH1 gene is reported in the majority of sporadic primary colorectal cancers with microsatellite instability, and that this methylation was often associated with loss of MLH1 protein expression (Herman et al. 1998).

Another epigenomic hallmark of cancer is the aberrant patterns of histone modifications. Epigenome-wide studies have characterized the overall profiles of various histone modification marks in cancer cells. For example, there is a global loss in H4K16ac in nearly all human cancer cell lines (Fraga et al. 2005b). Loss of acetylation is mediated by HDACs, which have been found to be overexpressed (Zhu et al. 2004) or mutated (Ropero et al. 2006) in different tumor types. Two different studies reported that global levels of H4K12ac and H3K18ac increased in adenocarcinomas in respect to normal tissue or adenoma (Ashktorab et al. 2009; Nakazawa et al. 2012). Cancer cells also bear global alterations of several histone methylation marks, such as a global loss of the active mark H3K4me3 (Hamamoto et al. 2004), and the repressive mark H4K20me3 (Fraga et al. 2005b), as well as a gain in the repressive marks H3K9me (Kondo et al. 2007) and H3K27me3 (Vire et al. 2006; Muller-Tidow et al. 2010).

Alterations of histone methylation marks in cancer cells are mainly due to the aberrant expression of both HMTs and histone demethylases (Chi et al. 2010). Gillian et al. reported the inactivating mutations in two genes encoding enzymes involved in histone modification: SETD2 gene (H3K36 methyltransferase) and JARID1C genes (H3K4 demethylase) in renal carcinomas (Dalgliesh et al. 2010). EZH2 gene (H3K27 methyltransferase) was reported overexpressed in several cancer types and enhances proliferation and neoplastic transformation (Kleer et al. 2003; Raman et al. 2005; Rhodes et al. 2004). NSD1, another HMT (H3K36 and H4K20), has been reported to undergo promoter DNA methylation-dependent silencing in neuroblastomas (Berdasco et al. 2009). H3K79 methyltransferase DOT1L is essential for development and maintenance the mixed lineage leukaemia. The presence of DOT1L results in H3K79 hypermethylation, which induces aberrant gene expression and contributes to leukemic transformation (Okada et al. 2006).

10.3.2 Epigenomics of Autoimmune Diseases

DNA methylation alteration has been increasingly associated with several autoimmune diseases in recent years; for which most studies focus on systemic autoimmune rheumatic diseases like systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA). SLE is characterized by autoantibody response to nuclear and/or cytoplasmic antigens. Several studies have shown that there is a global hypomethylation of promoter regions, which drive the genes that are overexpressed in the disease such as PRF1, CD70, CD154, IFGNR2, MMP14, LCN2, CSF3R and AIM2 genes, and also in the ribosomal RNA gene promoter (18S and 28S) (Portela and Esteller 2010; Ballestar 2011). This global loss of methylation has been attributed to induce the activation of endogenous retroviruses such that they erase imprinting signals and deregulate gene expression and consequently break immune tolerance for active flaring of the disease (Okada et al. 2002). The hypomethylation in SLE may be partially mediated by miR-21 and miR-148a that directly and indirectly target DNMT1 (Pan et al. 2010; Zhu et al. 2011). RA is a chronic inflammatory disease that largely affects peripheral joints by invasive synovial fibroblasts. Global changes in DNA methylation measured in fibroblast like synoviocytes showed distinct methylation profiles of RA patients, particularly in genes with key roles in inflammation, immune responses and matrix deconvolution. Hypomethylated loci were identified in key genes relevant to RA, such as CHI3L1, CASP1, STAT3, MAP3K5, MEFV and WISP3. Hypermethylation was also observed at some RA related genes, including TGFBR2 and FOXO1. Differentially methylated genes could alter fibroblast like synoviocytes gene expression and contribute to the pathogenesis of RA. Histone modification studies in human autoimmune diseases have found that during apoptosis, histones can be modified to make them immunogenic. Hypoacetylated histones H3 and H4 and H3K9 hypomethylation in $CD4^+$ T cells were found to be a characteristic feature of SLE patients (Hu et al. 2008). In RA, the reduced activity of HDACs plays a key role in regulating NF-κB–mediated gene expression (Huber et al. 2007).

10.4   Discussion and Perspectives

Advances in technological development have enabled epigenomic analysis on a large scale. Remarkably, several international projects and consortia (Table 10.2) have been formed to comprehensively characterize epigenome-wide DNA methylation, histone modification, and other epigenetic profiles in healthy and disease tissues, such as the Encyclopedia of DNA Elements (ENCODE) Project (Consortium 2012), the Cancer Genome Atlas (TCGA) (TCGA. The Cancer Genome Atlas. http://cancergenome.nih.gov/) and the NIH Roadmap Epigenomics Project (The NIH Roadmap Epigenomics Project, http://www.epigenomebrowser.org/).

Although the number of epigenomic studies has grown exponentially in recent years, several issues need to be carefully considered when planning and interpretation of such studies. First, disease-associated epigenetic variation is likely to be cell-/tissue-specific. For studies using heterogeneous cell/tissue samples (e. g. blood, tumor), detection of differential DNA methylation or histone modification profiles is a problem of validity: molecular profile variation and changes in cell type proportions between tissue samples are confounded (Jacobsen et al. 2006; Jaffe and Irizarry 2014). If the disease-associated variation is restricted to a certain cell type that represents only a small proportion of the tissue sampled, then the

Table 10.2 Large-scale national and international epigenomic consortia

Project name	Start date	Affiliations	Data contributions	Access data
The Encyclopedia of DNA Elements (ENCODE) Project	2003	NIH	ChIP-seq, RNA-seq, DNase-seq, shRNA knockdown followed by RNA-seq, RRBS, shotgun bisulfite-seq assay, DNA methylation profiling by array assay etc.in more than 200 of primary human tissues and cell lines	http://encodeproject.org/
The Cancer Genome Atlas (TCGA)	2006	NIH	Matched tumor and normal tissues from 11,000 patients, allowing for the comprehensive characterization of 33 cancer types and subtypes, including 10 rare cancers	http://cancergenome.nih.gov/
Roadmap Epigenomics Project	2008	NIH	Bisulfite-seq, MeDIP-seq, MRE-seq, RRBS, DNaseI, smRNA-seq, ChIP-seq etc. in more than 160s of normal primary cells, hESC, and hESC derived cells	http://www.epigenomebrowser.org/
International Cancer Genome Consortium (ICGC)	2008	17 countries, includes TCGA	DNA methylation profiles in thousands of patient samples from 31 tumor types	https://icgc.org/

variation may not be detected in the whole tissue (Jaffe and Irizarry 2014). Purified samples consisting only of a single cell type are preferable to mixed cell samples. Second, the complex system of the human body has many research areas, including genomics, epigenomics, transcriptomics, proteomics and metabolomics. Each research area provides insight into the system, but the entire complex of "omics" research offers more comprehensive insights. As costs of analysis of a human genome have dramatically plummeted, data integration is now a very commonly used notion. Integration between different epigenetic mechanisms and with other omics disciplines becomes easier and necessary for clinical research. For clinicians with access to omics data, being able to understand and appropriate interpret the data will become a key requirement for patient care. Along with the recent advancement in epigenetic drugs, there is a great potential for personalized epigenetic treatment of many human diseases in the near future.

References

Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012;13(10):R87. doi:10.1186/gb-2012-13-10-r87. PubMed PMID: 23034086; PubMed Central PMCID: PMC3491415.

Alegria-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. Epigenomics. 2011;3 (3):267–77. doi:10.2217/epi.11.22. PubMed PMID: 22122337, PubMed Central PMCID: PMC3752894.

Antelo M, Balaguer F, Shia J, Shen Y, Hur K, Moreira L, et al. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. PLoS One. 2012;7(9): e45357. doi:10.1371/journal.pone.0045357. PubMed PMID: 23049789, PubMed Central PMCID: PMC3458035.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363–9. doi:10.1093/bioinformatics/btu049. PubMed PMID: 24478339, PubMed Central PMCID: PMC4016708.

Ashktorab H, Belgrave K, Hosseinkhah F, Brim H, Nouraie M, Takkikto M, et al. Global histone H4 acetylation and HDAC2 expression in colon adenoma and carcinoma. Dig Dis Sci. 2009;54 (10):2109–17. doi:10.1007/s10620-008-0601-7. PubMed PMID: 19057998, PubMed Central PMCID: PMC2737733.

Ballestar E. Epigenetic alterations in autoimmune rheumatic diseases. Nat Rev Rheumatol. 2011;7 (5):263–71. doi:10.1038/nrrheum.2011.16. PubMed.

Barchitta M, Quattrocchi A, Maugeri A, Vinciguerra M, Agodi A. LINE-1 hypomethylation in blood and tissue samples as an epigenetic marker for cancer risk: a systematic review and meta-analysis. PLoS One. 2014;9(10):e109478. doi:10.1371/journal.pone.0109478. PubMed PMID: 25275447, PubMed Central PMCID: PMC4183594.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823–37. doi:10.1016/j.cell.2007.05.009. PubMed.

Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 2011;12(1):R10. doi:10.1186/gb-2011-12-1-r10. PubMed PMID: 21251332, PubMed Central PMCID: PMC3091299.

Berdasco M, Ropero S, Setien F, Fraga MF, Lapunzina P, Losson R, et al. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. Proc Natl Acad Sci U S A. 2009;106(51):21830–5. doi:10.1073/pnas.0906831106. PubMed PMID: 20018718, PubMed Central PMCID: PMC2793312.

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288–95. doi:10.1016/j.ygeno.2011.07.007. PubMed.

Bird AP. CpG-rich islands and the function of DNA methylation. Nature. 1986;321(6067):209–13. doi:10.1038/321209a0. PubMed PMID: WOS:A1986C330700035.

Bird AP, Southern EM. Use of restriction enzymes to study eukaryotic dna methylation .1. Methylation pattern in ribosomal dna from xenopus-laevis. J Mol Biol. 1978;118(1):27–47. doi:10.1016/0022-2836(78)90242-5. PubMed PMID: WOS:A1978EM86300002.

Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet. 2012;13(10):705–19. doi:10.1038/nrg3273. PubMed.

Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet. 2009;10(5):295–304. doi:10.1038/nrg2540. PubMed.

Cedar H, Solage A, Glaser G, Razin A. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. Nucleic Acids Res. 1979;6(6):2125–32. doi:10.1093/nar/6.6.2125. PubMed PMID: WOS:A1979GX25500006.

Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H, et al. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. Genome Res. 2010;20(10):1441–50. doi:10.1101/gr.110114.110. PubMed PMID: 20802089, PubMed Central PMCID: PMC2945193.

Chen PY, Cokus SJ, Pellegrini M. BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinf. 2010;11:203. doi:10.1186/1471-2105-11-203. PubMed PMID: 20416082, PubMed Central PMCID: PMC2871274.

Chen Z, Liu Q, Nadarajah S. A new statistical approach to detecting differentially methylated loci for case control Illumina array methylation data. Bioinformatics. 2012;28(8):1109–13. doi:10.1093/bioinformatics/bts093. PubMed PMID: 22368244, PubMed Central PMCID: PMC3324514.

Chi P, Allis CD, Wang GG. Covalent histone modifications–miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer. 2010;10(7):457–69. doi:10.1038/nrc2876. PubMed PMID: 20574448, PubMed Central PMCID: PMC3262678.

Cohen I, Poreba E, Kamieniarz K, Schneider R. Histone modifiers in cancer: friends or foes? Genes Cancer. 2011;2(6):631–47. doi:10.1177/1947601911417176. PubMed PMID: 21941619, PubMed Central PMCID: PMC3174261.

Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74. doi:10.1038/nature11247. PubMed PMID: 22955616, PubMed Central PMCID: PMC3439153.

da Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13. doi:10.1093/nar/gkn923. PubMed PMID: 19033363, PubMed Central PMCID: PMC2615629.

Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010;463 (7279):360–3. doi:10.1038/nature08672. PubMed PMID: 20054297, PubMed Central PMCID: PMC2820242.

Davis S DP, Bilke S, Triche JrT, Bootwalla M. Methylumi: handle illumina methylation data. R package version 220. 2012.

Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the infinium methylation 450K technology. Epigenomics. 2011;3(6):771–84. doi:10.2217/epi.11.105.

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol. 2008;26 (7):779–85. doi:10.1038/nbt1414. PubMed PMID: 18612301, PubMed Central PMCID: PMC2644410.

Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. Hum Mol Genet. 2007;16(Spec No 1):R50–9. doi:10.1093/hmg/ddm018. PubMed.

Esteller M. Epigenetics in cancer. N Engl J Med. 2008;358(11):1148–59. doi:10.1056/NEJMra072067. PubMed.

Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861–74. doi:10.1038/nrg3074. PubMed.

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A. 2005a;102 (30):10604–9. doi:10.1073/pnas.0500398102. PubMed PMID: 16009939, PubMed Central PMCID: PMC1174919.

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet. 2005b;37(4):391–400. doi:10.1038/ng1531. PubMed.

Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. J Mol Biol. 1987;196 (2):261–82. PubMed.

Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet. 2012;13(5):343–57. doi:10.1038/nrg3173. PubMed PMID: WOS:000303045400011.

Haines TR, Rodenhiser DI, Ainsworth PJ. Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. Dev Biol. 2001;240(2):585–98. doi:10.1006/dbio.2001.0504. PubMed PMID: WOS:000173216300021.

Halachev K, Bast H, Albrecht F, Lengauer T, Bock C. EpiExplorer: live exploration and global analysis of large epigenomic datasets. Genome Biol. 2012;13(10):R96. doi:10.1186/gb-2012-13-10-r96. PubMed PMID: 23034089, PubMed Central PMCID: PMC3491424.

Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, et al. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol. 2004;6 (8):731–40. doi:10.1038/ncb1151. PubMed.

Heerboth S, Lapinska K, Snyder N, Leary M, Rollinson S, Sarkar S. Use of epigenetic drugs in disease: an overview. Genet Epigenet. 2014;6:9–19. doi:10.4137/GEG.S12270. PubMed PMID: 25512710, PubMed Central PMCID: PMC4251063.

Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. Science. 2007;315(5815):1141–3. doi:10.1126/science.1136352. PubMed PMID: WOS:000244387600041.

Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci U S A. 1998;95(12):6870–5. PubMed PMID: 9618505, PubMed Central PMCID: PMC22665.

Hodawadekar SC, Marmorstein R. Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. Oncogene. 2007;26(37):5528–40. doi:10.1038/sj.onc.1210619. PubMed.

Hu N, Qiu X, Luo Y, Yuan J, Li Y, Lei W, et al. Abnormal histone modification patterns in lupus CD4+ T cells. J Rheumatol. 2008;35(5):804–10. PubMed.

Huber LC, Stanczyk J, Jungel A, Gay S. Epigenetics in inflammatory rheumatic diseases. Arthritis Rheum. 2007;56(11):3523–31. doi:10.1002/art.22948. PubMed.

Jacobsen M, Repsilber D, Gutschmidt A, Neher A, Feldmann K, Mollenkopf HJ, et al. Deconfounding microarray analysis – independent measurements of cell type proportions used in a regression model to resolve tissue heterogeneity bias. Methods Inf Med. 2006;45 (5):557–63. PubMed.

Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. 2014;15(2):R31. doi:10.1186/gb-2014-15-2-r31. PubMed PMID: 24495553, PubMed Central PMCID: PMC4053810.

Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–27. doi:10.1093/biostatistics/kxj037. PubMed.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996–1006. doi:10.1101/gr.229102. Article published online before print in May 2002. PubMed PMID: 12045153; PubMed Central PMCID: PMC186604.

Kidder BL, Hu G, Zhao K. ChIP-Seq: technical considerations for obtaining high-quality data. Nat Immunol. 2011;12(10):918–22. doi:10.1038/ni.2117. PubMed PMID: 21934668, PubMed Central PMCID: PMC3541830.

Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc Natl Acad Sci U S A. 2003;100(20):11606–11. doi:10.1073/pnas.1933744100. PubMed PMID: 14500907, PubMed Central PMCID: PMC208805.

Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472–9. doi:10.1038/nature12750. PubMed PMID: 24153300, PubMed Central PMCID: PMC4046508.

Kondo Y, Shen L, Suzuki S, Kurokawa T, Masuko K, Tanaka Y, et al. Alterations of DNA methylation and histone modifications contribute to gene silencing in hepatocellular carcinomas. Hepatol Res. 2007;37(11):974–83. doi:10.1111/j.1872-034X.2007.00141.x. PubMed.

Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571–2. doi:10.1093/bioinformatics/btr167. PubMed PMID: 21493656, PubMed Central PMCID: PMC3102221.

Ku CS, Naidoo N, Wu M, Soong R. Studying the epigenome using next generation sequencing. J Med Genet. 2011;48(11):721–30. doi:10.1136/jmedgenet-2011-100242. PubMed.

Kuan PF, Wang S, Zhou X, Chu H. A statistical framework for Illumina DNA methylation arrays. Bioinformatics. 2010;26(22):2849–55. doi:10.1093/bioinformatics/btq553. PubMed PMID: 20880956, PubMed Central PMCID: PMC3025715.

Kulis M, Esteller M. DNA methylation and cancer. Adv Genet. 2010;70:27–56. doi:10.1016/B978-0-12-380866-0.60002-2. PubMed.

Laird PW. Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet. 2010;11(3):191–203. doi:10.1038/nrg2732. PubMed.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. doi:10.1093/bioinformatics/btp324. PubMed PMID: 19451168, PubMed Central PMCID: PMC2705234.

Lim DH, Maher ER. Genomic imprinting syndromes and cancer. Adv Genet. 2010;70:145–75. doi:10.1016/B978-0-12-380866-0.60006-X. PubMed.

Lin X, Sun D, Rodriguez B, Zhao Q, Sun H, Zhang Y, et al. BSeQC: quality control of bisulfite sequencing experiments. Bioinformatics. 2013;29(24):3227–9. doi:10.1093/bioinformatics/btt548. PubMed PMID: 24064417, PubMed Central PMCID: PMC3842756.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462 (7271):315–22. doi:10.1038/nature08514. PubMed PMID: WOS:000271899300037.

Mancuso FM, Montfort M, Carreras A, Alibes A, Roma G. HumMeth27QCReport: an R package for quality control and primary analysis of Illumina Infinium methylation data. BMC Res Notes. 2011;4:546. doi:10.1186/1756-0500-4-546. PubMed PMID: 22182516, PubMed Central PMCID: PMC3285701.

Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerstrom-Billai F, Jagodic M, et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. Epigenetics. 2013;8(3):333–46. doi:10.4161/epi.24008. PubMed PMID: 23422812, PubMed Central PMCID: PMC3669124.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 2005;33(18):5868–77. doi:10.1093/nar/gki901. PubMed PMID: 16224102, PubMed Central PMCID: PMC1258174.

Micsinai M, Parisi F, Strino F, Asp P, Dynlacht BD, Kluger Y. Picking ChIP-seq peak detectors for analyzing chromatin modification experiments. Nucleic Acids Res. 2012;40(9):e70. doi:10.1093/nar/gks048. PubMed PMID: 22307239, PubMed Central PMCID: PMC3351193.

Muller-Tidow C, Klein HU, Hascher A, Isken F, Tickenbrock L, Thoennissen N, et al. Profiling of histone H3 lysine 9 trimethylation levels predicts transcription factor activity and survival in acute myeloid leukemia. Blood. 2010;116(18):3564–71. doi:10.1182/blood-2009-09-240978. PubMed PMID: 20498303, PubMed Central PMCID: PMC2981478.

Nagarajan RP, Zhang B, Bell RJ, Johnson BE, Olshen AB, Sundaram V, et al. Recurrent epimutations activate gene body promoters in primary glioblastoma. Genome Res. 2014;24 (5):761–74. doi:10.1101/gr.164707.113. PubMed PMID: 24709822, PubMed Central PMCID: PMC4009606.

Nakazawa T, Kondo T, Ma D, Niu D, Mochizuki K, Kawasaki T, et al. Global histone modification of histone H3 in colorectal cancer and its precursor lesions. Hum Pathol. 2012;43(6):834–42. doi:10.1016/j.humpath.2011.07.009. PubMed.

Ogino S, Nishihara R, Lochhead P, Imamura Y, Kuchiba A, Morikawa T, et al. Prospective study of family history and colorectal cancer risk by tumor LINE-1 methylation level. J Natl Cancer Inst. 2013;105(2):130–40. doi:10.1093/jnci/djs482. PubMed PMID: 23175808, PubMed Central PMCID: PMC3545905.

Okada M, Ogasawara H, Kaneko H, Hishikawa T, Sekigawa I, Hashimoto H, et al. Role of DNA methylation in transcription of human endogenous retrovirus in the pathogenesis of systemic lupus erythematosus. J Rheumatol. 2002;29(8):1678–82. PubMed.

Okada Y, Jiang Q, Lemieux M, Jeannotte L, Su L, Zhang Y. Leukaemic transformation by CALM-AF10 involves upregulation of Hoxa5 by hDOT1L. Nat Cell Biol. 2006;8(9):1017–24. doi:10.1038/ncb1464. PubMed PMID: 16921363, PubMed Central PMCID: PMC4425349.

Okano M, Xie SP, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998;19(3):219–20. PubMed PMID: WOS:000074565900012.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247–57. doi:10.1016/s0092-8674(00)81656-6. PubMed PMID: WOS:000083440600003.

Pan W, Zhu S, Yuan M, Cui H, Wang L, Luo X, et al. MicroRNA-21 and microRNA-148a contribute to DNA hypomethylation in lupus CD4+ T cells by directly and indirectly targeting DNA methyltransferase 1. J Immunol. 2010;184(12):6773–81. doi:10.4049/jimmunol.0904060. PubMed.

Park SY, Seo AN, Jung HY, Gwak JM, Jung N, Cho NY, et al. Alu and LINE-1 hypomethylation is associated with HER2 enriched subtype of breast cancer. PLoS One. 2014;9(6):e100429. doi:10.1371/journal.pone.0100429. PubMed PMID: 24971511, PubMed Central PMCID: PMC4074093.

Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 2012;7(2):e30619. doi:10.1371/journal.pone.0030619. PubMed PMID: 22312429, PubMed Central PMCID: PMC3270013.

Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. Nat Methods. 2009;6(11 Suppl):S22–32. doi:10.1038/nmeth.1371. PubMed PMID: 19844228, PubMed Central PMCID: PMC4121056.

Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010;28 (10):1057–68. doi:10.1038/nbt.1685. PubMed.

Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12(8):529–41. doi:10.1038/nrg3000. PubMed PMID: 21747404, PubMed Central PMCID: PMC3508712.

Raman JD, Mongan NP, Tickoo SK, Boorjian SA, Scherr DS, Gudas LJ. Increased expression of the polycomb group gene, EZH2, in transitional cell carcinoma of the bladder. Clin Cancer Res. 2005;11(24 Pt 1):8570–6. doi:10.1158/1078-0432.CCR-05-1047. PubMed.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A. 2000;97(10):5237–42. doi:10.1073/pnas.97.10.5237. PubMed PMID: WOS:000086998500044.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A. 2004;101(25):9309–14. doi:10.1073/pnas.0401994101. PubMed PMID: 15184677, PubMed Central PMCID: PMC438973.

Rivera CM, Ren B. Mapping human epigenomes. Cell. 2013;155(1):39–55. doi:10.1016/j.cell.2013.09.011. PubMed PMID: 24074860, PubMed Central PMCID: PMC3838898.

Robinson MD, Stirzaker C, Statham AL, Coolen MW, Song JZ, Nair SS, et al. Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. Genome Res. 2010;20(12):1719–29. doi:10.1101/gr.110601.110. PubMed PMID: 21045081, PubMed Central PMCID: PMC2989998.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6. doi:10.1038/nbt.1754. PubMed PMID: 21221095, PubMed Central PMCID: PMC3346182.

Ropero S, Fraga MF, Ballestar E, Hamelin R, Yamamoto H, Boix-Chornet M, et al. A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition. Nat Genet. 2006;38(5):566–9. doi:10.1038/ng1773. PubMed.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. doi:10.1101/gr.1239303. PubMed PMID: 14597658, PubMed Central PMCID: PMC403769.

Siegmund KD. Statistical approaches for the analysis of DNA methylation microarray data. Hum Genet. 2011;129(6):585–95. doi:10.1007/s00439-011-0993-x. PubMed PMID: 21519831, PubMed Central PMCID: PMC3166559.

Sterner DE, Berger SL. Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev MMBR. 2000;64(2):435–59. PubMed PMID: 10839822, PubMed Central PMCID: PMC98999.

Sun Z, Chai HS, Wu Y, White WM, Donkena KV, Klein CJ, et al. Batch effect correction for genome-wide methylation data with Illumina Infinium platform. BMC Med Genomics. 2011;4:84. doi:10.1186/1755-8794-4-84. PubMed PMID: 22171553, PubMed Central PMCID: PMC3265417.

Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465–76. doi:10.1038/nrg2341. PubMed.

Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012;4(3):325–41. doi:10.2217/epi.12.21. PubMed.

van Hoesel AQ, van de Velde CJ, Kuppen PJ, Liefers GJ, Putter H, Sato Y, et al. Hypomethylation of LINE-1 in primary tumor has poor prognosis in young breast cancer patients: a retrospective cohort study. Breast Cancer Res Treat. 2012;134(3):1103–14. doi:10.1007/s10549-012-2038-0. PubMed.

Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, et al. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 2013;23(3):555–67. doi:10.1101/gr.147942.112. PubMed PMID: 23325432, PubMed Central PMCID: PMC3589544.

Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, et al. The Polycomb group protein EZH2 directly controls DNA methylation. Nature. 2006;439(7078):871–4. doi:10.1038/nature04431. PubMed.

Waddington CH. Canalization of development and genetic assimilation of acquired characters. Nature. 1959;183(4676):1654–5. PubMed.

Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol. 2014;15(2):R37. doi:10.1186/gb-2014-15-2-r37. PubMed PMID: 24555846, PubMed Central PMCID: PMC4053980.

Wang S. Method to detect differentially methylated loci with case-control designs using Illumina arrays. Genet Epidemiol. 2011;35(7):686–94. doi:10.1002/gepi.20619. PubMed PMID: 21818777, PubMed Central PMCID: PMC3197755.

Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, et al. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics. 2012;28(5):729–30. doi:10.1093/bioinformatics/bts013. PubMed PMID: 22253290, PubMed Central PMCID: PMC3289916.

Wilbanks EG, Facciotti MT. Evaluation of algorithm performance in ChIP-seq peak detection. PLoS One. 2010;5(7):e11471. doi:10.1371/journal.pone.0011471. PubMed PMID: 20628599, PubMed Central PMCID: PMC2900203.

Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, et al. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet. 2010;6(4):e1000917. doi:10.1371/journal.pgen.1000917. PubMed PMID: 20421991, PubMed Central PMCID: PMC2858672.

Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol. 2010;11(9):607–20. doi:10.1038/nrm2950. PubMed PMID: 20683471, PubMed Central PMCID: PMC3711520.

Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinf. 2009;10:232. doi:10.1186/1471-2105-10-232. PubMed PMID: 19635165, PubMed Central PMCID: PMC2724425.

Xi Y, Bock C, Muller F, Sun D, Meissner A, Li W. RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing. Bioinformatics. 2012;28 (3):430–2. doi:10.1093/bioinformatics/btr668. PubMed PMID: 22155871, PubMed Central PMCID: PMC3268241.

Zhu P, Martin E, Mengwasser J, Schlag P, Janssen KP, Gottlicher M. Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis. Cancer Cell. 2004;5(5):455–63. PubMed.

Zhu X, Liang J, Li F, Yang Y, Xiang L, Xu J. Analysis of associations between the patterns of global DNA hypomethylation and expression of DNA methyltransferase in patients with systemic lupus erythematosus. Int J Dermatol. 2011;50(6):697–704. doi:10.1111/j.1365-4632.2010.04804.x. PubMed.