Integrating Epigenomics into the Understanding of Biomedical Insight

Yixing Han[1,2] and Ximiao He[3,4]

[1]Mouse Cancer Genetics Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD, USA. [2]Present address: Genetics and Biochemistry Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [3]Laboratory of Metabolism, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [4]Present address: Department of Medical Genetics, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China.

ABSTRACT: Epigenetics is one of the most rapidly expanding fields in biomedical research, and the popularity of the high-throughput next-generation sequencing (NGS) highlights the accelerating speed of epigenomics discovery over the past decade. Epigenetics studies the heritable phenotypes resulting from chromatin changes but without alteration on DNA sequence. Epigenetic factors and their interactive network regulate almost all of the fundamental biological procedures, and incorrect epigenetic information may lead to complex diseases. A comprehensive understanding of epigenetic mechanisms, their interactions, and alterations in health and diseases genome widely has become a priority in biological research. Bioinformatics is expected to make a remarkable contribution for this purpose, especially in processing and interpreting the large-scale NGS datasets. In this review, we introduce the epigenetics pioneering achievements in health status and complex diseases; next, we give a systematic review of the epigenomics data generation, summarize public resources and integrative analysis approaches, and finally outline the challenges and future directions in computational epigenomics.

KEYWORDS: epigenetics, computational epigenomics, chromatin, DNA methylation, histone modification, ncRNAs, NGS, integrative analysis

CITATION: Han and He. Integrating Epigenomics into the Understanding of Biomedical Insight. *Bioinformatics and Biology Insights* 2016:10 267–289 doi: 10.4137/BBI.S38427.

TYPE: Review

RECEIVED: June 17, 2016. **RESUBMITTED:** November 01, 2016. **ACCEPTED FOR PUBLICATION:** November 06, 2016.

ACADEMIC EDITOR: Thomas Dandekar, Associate Editor

PEER REVIEW: Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 573 words, excluding any confidential comments to the academic editor.

FUNDING: The Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research supported this work. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: yi-xing.han@nih.gov

Introduction

With the advance of the next-generation sequencing (NGS) technology, large-scale omics data are accumulating at an exponential growth rate. It drives the biomedical study and the understanding of the life science to be increasingly data intensive. Scientific discoveries are based more and more on the genome-wide scale data and systematic data analysis. However, genome research is still facing significant challenges, including the shifts of the bottleneck from data generation to data analysis and data interpretation and aggravation of the difficulty of the integrative analysis in dimensions.

The field of epigenetics and epigenomics is attracting immense interest with countless studies. Epigenetics is defined as the "stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence".[1] Epigenetic regulation comprises many different pathways such as DNA methylation, histone modifications, histone variants, nucleosome positioning, and noncoding RNAs (ncRNAs). These factors work on the interface of the environment and the genome and play an essential role in fundamental biological processes, which touch upon the main central problems of biology: How do the epigenetic mechanisms work as a driving force in the cell specialization during development[2]? Which molecular mechanisms contribute to phenotypic inheritance and evolutionary adaptation[3,4]? And how epigenetic factors influence the complex diseases[4–6]?

Different categories of epigenetic regulatory factors are involved in an interactive network and act coordinately within or between chromosomes to shape the genomic architecture, regulation, and transcriptional and translational outcomes. Epigenomics extends the epigenetics study from locus and single factors to global and multiple layers of regulatory cues. It is essential studies for the landscapes establishment of epigenetic marks under various conditions, which facilitates the understanding that the epigenetic profiles are maintained and affected via machinery that is regulated by the cross talk among these layers and the interplay with binding proteins, chromatin accessibility, and 3D conformation.[7,8] From the genome and interaction network points of view, NGS was widely adopted promptly after its development in this field and generated comprehensive massive genome-wide datasets in all the epigenetic regulation layers.[9] Hence, the joint analysis of multilayer epigenomic data, together with genomic, transcriptomic, and proteomic data through integration methods, is

critical to comprehend how epigenetic information contributes controlling complex regulatory processes.

Here, we review pioneering epigenomic studies and computational analyses that have contributed to biomedical research. In addition, we summarize the data, tools, and resources and outline future challenges in computational epigenetics that is super valued in addressing the full picture of the biological system.

Epigenetic Mechanisms

In eukaryotic cells, genomic DNA is compacted more than 10,000-fold in the nucleus by wrapping around highly conserved proteins termed as histones. This highly assembled DNA–protein structure is called nucleosome that forms the building blocks of chromatin. In general, the tighter the DNA is wrapped up, the more likely the gene is repressively expressed, while more accessible chromatin (less condense chromatin structure) indicates that the transcription machinery will be easy to bind and start up the gene transcription. It is the covalent modification that the epigenetic inheritance is encoded in, rather than the DNA sequence (which is the genetic inheritance encoded in). Epigenetic information can faithfully propagate between generations of cells (mitotic inheritance)[2] and between generations of species (meiotic inheritance),[10] but with substantially lower fidelity than genetic information.[11,12]

There are four types of epigenetic regulators: DNA methylation, histone modification, nonhistone binding proteins, and ncRNAs that act synergistically to control the chromatin architecture for cellular processes such as transcription, replication, and DNA repair.[13] DNA is subject to be methylated at specific regions, so that it can foster a locally more compact chromatin structure and influence the accessibility for transcription factors.[14] The histones consist of four core histones (two copies of H2A, H2B, H3, and H4) that are subject to a large number of posttranslational modifications on the unstructured N-terminal tails, including lysine and arginine methylation, lysine acetylation, and serine phosphorylation.[15] Moreover, nonhistone proteins can affect the chromatin structure by interacting with histone and DNA in a variety of ways. ATP-dependent chromatin remodeling factors can directly mobilize nucleosomes or work together with enzymes in DNA methylation pattern determination and histone code programming.[16–19] Epigenetic modifiers can dynamically "write", "read", and "erase" modifications to program/reprogram the chromatin accessibility to regulate gene expression during cell differentiation and disease occurrence.[20,21] They work jointly with DNA, histone, and nonhistone proteins to form a complex interaction network in regulating chromatin accessibility for transcription.[22] ncRNAs are RNAs that are transcribed from DNA but function as structural, functional, and regulatory molecules rather than serving as templates for proteins, which take up to 70% of the genome.[23–26] Based on the length of the ncRNAs and the biogenesis procedure, the epigenetic-related ncRNAs can be grouped into long noncoding RNAs (lncRNAs) (>200 nt), mid-size RNAs (20–300 nt), which include small nucleolar RNAs (snoRNAs),[27,28] promoter-associated small RNAs (PASRs), TSS-associated RNAs (TSSa-RNAs),[29] and short ncRNAs (<200 nt), which include microRNAs (miRNAs; 21–23 nt),[30,31] short interfering RNAs (siRNAs; 20–30 nt),[31] Piwi-interacting RNAs (piRNAs; 27–30 nt),[32] and tRNA-derived RNAs (tDRs; 20–35 nt).[33,34] The mechanism by which a vast void of these ncRNAs function and process remains to be discovered; however, well-studied cases show that these ncRNAs can interact with DNA, RNA, and proteins and generally function as *cis*-acting silencers and also *trans*-acting mediators for site-specific transcriptional and posttranscriptional processes, nuclear organization, RNA processing, and transposon suppression through sequence complementary.[35–38] The study of mid-size ncRNAs and lncRNA is still in its infancy, and their biological functions are predicted to be transcription relevant but remain to be well defined. However, several possible mechanisms for lncRNA have been proposed based on the few relatively well-studied examples. It has been uncovered that lncRNAs can form complexes with other factors against *cis*- (eg, enhancer-like activities) or *trans*-targets (eg, Hox transcript antisense intergenic RNA [HOTAIR] binding with polycomb repressive complex) and function both in nuclear and cytoplasm to regulate transcription and translation.[39–46] Thus, epigenetic mechanisms are fundamental to the regulation of many cellular processes, including the spatial and temporal expressions of gene and ncRNA, cell differentiation, embryogenesis, DNA replication, DNA repair, alternative splicing, X-chromosome inactivation, genome imprinting, and suppression of transposable element mobility.[23,47–53]

Beyond the epigenetic modifications occurring at linear chromatin domains, higher order chromatin territories are emerging with NGS technology as an important regulator of genes, which confirmed the findings by microscopy studies decades ago that chromosomes are positioned with preferential spatial in a nucleus to facilitate necessary long-range domain interaction and regulation.[54] Interactome studies revealed that the boundaries of topological domains are highly conserved across species and enriched for essential genes, repeat elements, and insulator-binding motifs.[55] Histone modification patterns can also be identified at topologically associating domains.[56] Active chromatin reorganizations occur in and regulate extensive biological procedures, including cell differentiation and tumorigenesis,[57,58] and have been found playing more and more instrumental roles in the epigenomics network.

Epigenomics Complex Diseases

The faithful propagation of epigenetic information is as important as the genetic information, which ensures the precise regulation of biological process over multiple cell divisions. Stochastic and environment-induced epigenetic defects are known to play a major role in occurrence of complex diseases,

including cancer, aging, mental disorders,[59] and autoimmune diseases.[60] Epigenetic mutations accumulate along with age and may result in nonproper activation of normally downregulated genes,[61,62] affecting genome stability.[63] These changes underlie general effects of aging and aging-related diseases, like cancer and neurodegenerative diseases.[59,64] For instance, DNA methylation pattern, due to the delicate balance between stability and plasticity, has been suggested to provide a lifetime record of environmental exposures and a valuable biomarker for risk stratification and disease diagnosis. Monozygotic twins, as they age, exhibit remarkable difference in genome methylation patterns that result in differential gene expression and, ultimately, life span.[65] A global demethylation occurred in DNA repeat elements in cancer and aging,[66,67] while the cancer epigenome is characterized by a massive global loss of DNA methylation[68] and a certain promoter CpG islands hypermethylation[69] that frequently overlaps with enhancers and other regulatory elements. The genome-wide DNA methylation changes mediate genome instability, chromosomal translocations, gene mutations, and reactivation of endoparasitic sequences. DNA methyltransferases (DNMTs) have been identified in elevated expression during aging and tumorigenesis, which are responsible for the hypomethylation feature.[70–72] Accumulating evidence supports the notion that DNA methylation constitutes a promising and reliable biomarker in clinical practice for earlier and more reliable cancer diagnosis[73,74] and more precise tumor subtype classification.[75,76]

Global profile changes of histone modifications and chromatin-modifying enzymes expression are also critical in aging[62,77] and cancer initiation and progression.[67,78] For example, cancer cells suffer a global reduction of activation markers H4K16ac[79] and H3K4me3[80] and a gain in the repressive markers H4K20me3,[79] H3K9me3,[81] and H3K27me3,[82] while H4K16ac,[83] H3K4me2,[62] H3K4me3,[84] and H4K20me3[85] are increased with age. Distribution alteration of the histone modifications is mainly due to the abnormal expression of histone-modifying enzymes, such as histone deacetylases (HDACs) in the sirtuin family,[86] SETD2,[87] and EZH2,[88] which lead to nonadaptive alterations of epigenetic landscape, thereby gene expression change. Deregulated epigenetic mediators that lead to the complex disorders may serve as potential targets of therapeutics, termed as epigenetic therapy. A great example is the sirtuin family of protein deacetylases that can be used as target to extend the health span of life. Small molecules that increase nicotinamide adenine dinucleotide phosphate (NADP) level can activate the sirtuins to mimic the effect of caloric restriction on genome-wide gene expression,[89,90] so that it represents an epigenetic interventional path to prevent neurodegeneration,[91] type II diabetes,[92] cancer,[93] and aging.[94]

ncRNAs are instrumental regulatory elements for cellular homeostasis. Rapidly growing evidences have consistently proved that the deregulation in their precise transcription and maturation, correct interaction with target mRNAs, and mutations in the ncRNA-processing machinery are causal factors in neurological, tumor genesis, and cardiovascular and developmental diseases. Among the variety of ncRNAs, miRNA is the most thoroughly studied one, especially in cancer and neurological disorders. miRNA can serve as both oncogenes and tumor suppressors, and the expression profiles are different between tumor and normal tissues and also among different cancer types,[95,96] which provide important information for cancer prognosis and classification. For example, the dysregulation of miR-15, miR-16,[97] and miR-200[98] family is associated with genetic alterations that affect their primary processing, maturation, and interaction with mRNA targets and leads to chronic lymphocytic leukemia (CLL), ovarian cancer, and breast cancer. The impairs in miRNA processing complexes that cause the abnormal maturation are involved in cancer, for example, mutations on TARBP2[99] and DICER1,[100] which are key processors in primary miRNA maturation, can cause downregulation of miRNA and then tumor genesis. It has been documented that approximately 70% of miRNAs are expressed in brain and specifically function in neural differentiation, maintenance, and synapsis plasticity, and their dysregulation has been found in almost all of the neurological disorders. For example, miR-29,[101] miR-107,[102] miR-298, and miR-328[103] regulate the beta-amyloid precursor protein-cleaving enzyme I and can accelerate Alzheimer's disease progression. Mutations on miRNA-processing machinery factors and RNA-binding proteins (RBPs) such as fragile X mental retardation 1 protein (FMRP) in RISC complex can cause fragile X syndrome (FXS),[104] lucine-rich repeat serine/threonine-protein kinase 2 (LRRK2) is a cause of Parkinson's disease,[105] and RBP Musashi1 is associated with many cancers, including breast, colon, glioblastoma, and medulloblastoma, as well as neurodegenerative diseases.[106] Numerous evidences are rapidly increasing about the lncRNA dysregulation in diseases, such as HOTAIR[107] and lincRNA-p21[108] in cancer and H19 in Silver–Russell syndrome and Beckwith–Wiedemann syndrome.[109]

Further understandings of the global patterns of these epigenetic modifications and their corresponding changes in complex diseases have enabled the diagnosis improvement, therapy target discovery, and better treatment strategy design.

Epigenomics Data Generation

Different approaches have been developed to capture the multiple levels of epigenetic signal for finally disentangling the epigenetic regulation network. Actually, most of these approaches follow a three-phase strategy. First, epigenetic information is converted into genetic information through biochemical methods. Next, standard DNA array technology or high-throughput sequencing is applied. Finally, computational and statistical analyses are then used to extract the sequence and infer the outcomes for biological insight interpretation.

With the combination of experiment and high-throughput sequencing technology, we have been able to acquire the data at large genomic regions and even genome-wide scale. Various experimental methods have been developed to identify DNA methylation patterns. Pretreatments for these methods use endonuclease digestion (such as CHARM[110] and MCA[111]), affinity enrichment (such as MeDIP[112] and MIRA[113]), and bisulfite conversion.[114] With the advances in NGS technologies, bisulfite conversion of unmethylated Cs to Ts followed by high-throughput sequencing (BS-seq) is a golden standard method to study the methylation status of every cytosine in the genome and produce the detailed DNA methylation maps.[115] Among them, the popular strategies include Whole-Genome Bisulfite Sequencing (WGBS)[116] and Reduced Representation Bisulfite Sequencing (RRBS).[117] These popular new methods produce huge volume of DNA methylation datasets (from hundred gigabytes to terabytes), which pose enormous challenges in terms of computational approaches to analyze and interpret these data.[114,115]

The histone modification signals and chromatin-binding factors can be captured by chromatin immunoprecipitation (ChIP)-based techniques, such as ChIP-seq[118,119] and ChIP-chip,[120] in which specific antibodies were used to enrich the DNA fragments at modification sites. The ultra high throughput flow approaches are becoming more and more popular due to its high coverage, high resolution, and low cost.[118,119,121,122] At specific region of the genome, chromatin has lost its condensed structure and exposed the DNA and makes it accessible for DNA degradation enzymes such as DNase I and transcriptional machinery. DNase-seq[123] utilizes the dynamic DNase I hypersensitive sites (DHSs) and combines the NGS to understand the chromatin package under various circumstances.

Recently, chromosome conformation capture (3C)-based techniques have been used increasingly to facilitate the detection of genome folding, chromosome spatial conformation, and long-range gene–gene interaction.[124] Particularly, an advanced 3C – Hi-C has been developed as a powerful tool for genome-wide intra- and interchromosomal interplay, which provides unbiased large-scale information for reconstruction of the 3D structure of the chromosome.[125] Furthermore, single-cell Hi-C significantly promoted the discovery of cell-to-cell variability in chromosome structure under normal cell status and disease conditions.[126,127]

Increasing novel classes of ncRNA are emerging from the 90% transcribed genome[25] with the application of NGS, which offers unprecedented opportunity to obtain higher throughput and accuracy and lower experimental complexity. The discovery and detection of the ncRNAs are mostly based on the size fractionation methods in the isolated RNAs that led to the identified classes of ncRNAs as small, mid-size, and long. Small RNA-seq is the popularly used approach for small ncRNA identification; the library construction has a large overlap with the RNA-seq with ribosome RNA elimination, cDNA synthesis, 3′-A addition, adaptor ligation, and PCR enrichment.[128] However, precise size selection of 18–30 nt or 30–200 nt fragments instead of the RNA fragmentation step is critical.[129] Due to the poly(A) tail and mRNA-like features, lncRNAs are able to be detected in the cDNA cloning, tilling array, and polyadenylated transcriptome data. For example, cDNA cloning followed by Sanger sequencing was used in the first large-scale (>34,000) lncRNAs cataloging from the FANTOM project[130,131] and in the lncRNA annotation from RefSeq and Ensembl projects.[132,133] Genome-wide tiling array of transcriptome also contribute more efficiently in the ncRNA identification,[26] and recently, high-throughput RNA-seq promotes the discovery sensitive dramatically and enables the reconstruction of the transcript models with or without a reference genome.[134,135]

The question that how the ncRNAs interact with DNA, mRNA, and proteins is in the central place of the ncRNA epigenetic regulation and functional annotation studies. Experimental approaches that were used for mRNA detection and quantification such as qPCR, Northern blots,[136] fluorescence in situ hybridization (FISH),[137] and RNA interference (RNAi)[138,139] can also be applied to the characterization of ncRNAs. RNA-binding protein immunoprecipitation (RIP)[140] followed by chip[141] or NGS sequencing[142] and UV cross-linking and immunoprecipitation (CLIP)[143] enable various RNA–RBP interaction studies with lower background and higher affinity. The application of CLIP-seq is expanding from mRNA to miRNA,[144] lncRNA,[145] cirRNA,[146] and mitochondrial RNA.[147] The genome-wide CLIP experiments should be designed specifically to accommodate the different aims of each study, for example, studies may focus on RBP-binding site identification, RBP interactions with other factors, and RBP function in different biological processes including transcription, splicing, and translation. Furthermore, chromatin isolation by RNA purification (ChIRP)-seq can apply to illuminate the interaction of RNA, chromatin, and protein.[148]

We summarized these approaches in Table 1. These rapidly advancing technologies create ample opportunities for epigenome research; however, at the meantime, they also pose substantial challenges in terms of large datasets' storage and processing, statistical analysis, and biological interpretation for observed differences.

Epigenomics Data Analysis

DNA methylations. In vertebrates, the most common form of DNA methylation is 5-methylcytosine (5-mC), which mainly occurs in the sequence context of CG dinucleotides. The non-CG methylation in a CHG or CHH context (where H stands for A, C, or T) exists in embryonic stem cells,[116] brain,[149,150] and plant.[151] In mammalian genomes, CG dinucleotides are rare but tend to occur in clusters called CG islands (CGI) that are often located in the proximal promoters of genes, particularly housekeeping genes,[152–154] but are typically not methylated. In the early embryo, there is little CG

Table 1. Main epigenomics data generation methods.

APPLICATION	METHODS	PRINCIPLE	REFS
DNA methylation pattern detection	Methylated DNA immunoprecipitation (MeDIP)	Purified DNA is immunoprecipitated with an antibody against methylated cytosines, giving rise to genomic maps of DNA methylation	111
	Bisulfite sequencing	Bisulfite to convert the unmethylated cytosines to uracils	114
	Reduced representation bisulfite sequencing (RRBS)	Combines restriction enzymes and bisulfite sequencing in order to enrich for the areas of the genome that have a high CpG content	116
Histone modification patter detection, chromatin binding protein pattern detection	ChIP chip	Specific antibodies used for enrichment of the DNA fragments at modification sites followed by array hybridazation	119
	ChIP-seq	Specific antibodies used for enrichment of the DNA fragments at modification sites followed by high-throughput sequencing	117,118
3D structure of chromatin	DNase-seq	At Dnase I hypersensitive sites (DHSs), chromatin are sensitive to cleavage by the Dnase I enzyme. These accessible chromatin zones are functionally related to transcriptional activity	122
	Hi-C chromosome conformation capturing technique	Chromosome contacts are captured by formaldehyde cross-linking	124,125,127
RNA-protein and RNA-DNA interaction	RIP-chip	Specific antibodies used for immunopreciptation of the RNA fragments at RNA-binding sites followed by reverse transcription and microarray	141
	RIP-seq	Specific antibodies used for immunopreciptation of the RNA fragments at RNA-binding sites followed by reverse transcription and high-throughput sequencing	142
	CLIP-seq	UV cross-linking with immunoprecipitation to analyze protein interactions with RNA to precisely locate RNA-protein binding site and RNA modifications. Modified versions including PAR-CLIP (photoactivatable-ribonucleoside-enhanced CLIP) can improve the signal-to-noise ratio and iCLIP (Individual-nucleotide resolution CLIP) can achieve a higher efficiency in reverse-transcription.	143,321,322
	ChIRP-seq	Biotin labeled oligos that are complement to interested RNA are used to hybridize crosslinked chromatin fragments to capture biotin-oligo-RNA-DNA-protein complexes, DNA then isolated from the complexes for high-throughput sequencing to illustrate the RNA-DNA interaction	148

methylation, but CG dinucleotides outside of CGI typically become methylated during the blastula stage of development.[14] It is mainly CG-rich regions outside of proximal promoters that become demethylated upon cellular differentiation.[117,155] However, genomic analyses have identified low CG promoters that are both methylated and transcriptionally active.[156–158]

Since the principles, computational methods, and challenges of DNA methylation have been heavily reviewed,[114,115,159] this review aims to put a particular emphasis on the computational approaches to BS-seq data (WGBS and RRBS), including essential steps of mapping BS-seq reads to the reference genome, determining DNA methylation level, detecting the differentially methylated regions (DMRs) between cases and controls, as well as storing, retrieving, and visualizing DNA methylation data.

Mapping BS-seq reads. Bisulfite treatment of DNA followed by PCR amplification and then sequencing leads to the vast majority of unmethylated Cs that are changed to Ts in the sequencing reads, without affecting As, Gs, Ts, or methylated Cs. To calculate the absolute DNA methylation level for each C from BS-seq data, the sequencing reads are required to align to the reference genome to determine the position where the reads were most likely to be derived. Various alignment tools, including the general aligners with BS-seq module and the specific BS-seq aligners, have been developed to map the BS-seq short reads (Table 2). Due to the specificity of the BS-seq reads, some general aligners are developed with BS-seq modules (such as GSNAP,[160] LAST,[161] Novoalign,[162] RMAP,[163] and segemehl[164]). Specific BS-seq aligners were also developed to map the BS-seq reads. Among these tools, two alternative approaches have been widely used. The three-letter aligners (such as Bismark,[165] BRAT,[166] BS-Seeker,[167] and MethylCoder[168]) simplify the alignment by converting all Cs into Ts for the BS-seq reads and both strands of

the reference genome (only three alphabets of A, G, and T remaining in the converted sequences) then using the standard aligner. In contrast, the wild-card aligners (such as BSMAP,[169] Pash,[170–172] and RRBSMAP[173]) only convert Cs to the wild-card letter Y (stands for pyrimidine: C or T), which matches both Cs and Ts in the BS-seq reads. The three-letter aligners reduce the sequence complexity, resulting in a higher percentage of discard reads owing to multiple alignments in the reference genome, while the wild-card aligner can achieve a higher genomic coverage but with some bias toward increased DNA methylation level.[115] After the alignment, the DNA methylation level can be determined by comparing the frequency of Cs and Ts that align to each C in the reference genome.

Detecting DMRs. After BS-seq reads mapping, the next step is typically the detection of DMRs that show significantly different DNA methylation levels between sample groups, such as disease versus normal, or cases versus controls. Based on the biological question of interest and different computational approaches of identification, these DMRs can range in size from as small as a single C site (differentially methylated C site [DMC]) to as large as an entire gene locus with length of megabase pairs. The most common methods to detect DMRs involve testing single C to identifying the DMCs by different statistical analysis and merging the significant DMCs into DMRs using various approaches.[115] The basic statistical tests for comparing the DNA methylation levels of each C with sufficient pooled data between sample groups are *t*-test, Wilcoxon rank-sum test, or linear regression.[115,174] Some more advanced models have been employed to improve the DMR detection, including beta regression and hierarchical testing (BiSeq[175]), weighted generalized linear model (BSmooth[176]), bump hunting with batch effect removal and peak detection (bumphunting[177]), tunable kernel smoothing (DMRcate[178]), nonparametric and kernel-based method (M3D[179]), beta-binomial model (methylSig[180]), beta-binomial hierarchical model (MOABS[181]), hidden Markov model (NHMMfdr[182]), three-state HMM (MethPipe[183]), Shannon entropy (QDMR[184]), and a binary segmentation algorithm combined with a two-dimensional statistical test (metilene[185]). Usually, the latest software compares with some previous methods and claims best performance, such as MOABS, which can detect the DMRs with a relative low coverage (~10×)[181] and metilene can identify DMRs with unrivaled specificity and sensitivity.[185] However, without a systematic benchmarking study, it is difficult to determine which methods will work best for the DNA methylation datasets. To address this issue, it is necessary to carry out the comprehensive comparison between these different DMR callers.

Currently, there are still some limitations for the BS-seq technology, such as its inability to distinguish between 5-mC and 5-hmC (5-hydroxymethylcytosine), and single-cell DNA methylation profile is yet to be developed. To overcome these limitations, the technologies such as oxidative bisulfite sequencing (oxBS-seq)[186] and Tet-assisted bisulfite sequencing (TAB-seq)[187] to distinguish 5-hmC from 5-mC, single-cell bisulfite sequencing using RRBS[188] or PBAT (post-bisulfite adaptor tagging),[189] and technologies enabling direct detection of modified bases (5-mC or 5-hmC) within individual DNA[190–192] have been introduced. With these new technologies, more elaborate and powerful bioinformatics software as well as web-based tools and resources will be developed, which is another great opportunity for computational epigenomics.

Major Bioinformatics Challenges in Interpreting DNA Methylation Differences

There are some major bioinformatics challenges in downstream interpretation of DNA methylation differences after DMR detecting and DNA methylation data visualizing. First of all, as mentioned in the detecting DMR section, it is difficult to compare the different methods without knowing the true methylation status in a certain biological sample. Second, it is more complicated when considering the variation of biological samples.[115] There are four major different levels of variations: (1) allele-specific DNA methylation is widespread even in the same cell, and some bioinformatic methods have been introduced to identify the DNA methylation differences between alleles[193,194]; (2) age-related and interindividual differences in DNA methylation is common and may be influenced by genetic differences[195–198]; (3) cell-specific methylation is observed in different cell types in the same tissue or organ[199,200]; and (4) the most complicated case is cancer sample, which is a mixture of tumor and normal cells with increased methylation variations.[201] Several bioinformatic tools have been developed to estimate the tumor purity.[202,203] Third, the most challenging computational analysis is correlating the DNA methylation differences with diseases. The challenges include the following: (1) the correlation of DNA methylation in promoter and gene expression is modest[200,201]; (2) the methylation changes can occur not only in promoter regions but also in other genic and intergenic regions[204]; and (3) the correlation does not necessarily mean causation. However, epigenome-wide association study (EWAS) has been introduced to identify the loci with DNA methylation variation, which is associated with common diseases.[205] Abnormal DNA methylation status (either in a CpG-rich region or a single CpG site) has been heavily studied as potential biomarkers for different cancer types,[206] such as colon cancer,[207–209] prostate cancer,[210–212] and lung cancer.[213,214]

Histone Modifications and DNA-Binding Proteins

The modifications on the unstructured histone tails control the accessibility of the chromatin for the transcription machinery as actively transcribed euchromatin or transcriptionally inactive heterochromatin. Euchromatin is characterized by high levels of acetylation and trimethylated H3K4, H3K36, and H3K79, while heterochromatin is characterized by low levels of acetylation and high levels of methylation on

H3K9, H3K27, and H4K20.[215] DNA-binding proteins bind preferentially to certain DNA sequences (termed as motif) and work together with histone marks to carry out cellular functions. Evidences are accumulating that the gene expression is predictable by the key factors binding, the histone modification levels, and the cross talks among the different modifications that occurred on the histone simultaneously,[216,217] for example, the active mark H3K4me3 and repressive mark H3K27me3 occupied "bivalent domain", which is pivotal for the embryonic stem cells (ESCs) pluripotent and differentiation states determination.[218]

ChIP followed by microarray or sequencing has become the widely used technique for identifying the histone modification and protein-binding locations and patterns genome widely.[219,220] Moreover, there are various adaptations of the standard ChIP protocol to overcome the limitations for a certain specific application. For instance, in order to use limited cells instead of the conventional 10 million cells for one ChIP reaction, Nano-ChIP-seq for H3K4me3 in 10,000 cells[221] and single-tube linear DNA amplification (LinDA) for Erα in 5,000 cells[222] have been successfully applied. ChIP-exo using the lambda phage exonuclease feature is able to remarkably enhance the binding precision to single base pair and significantly decrease the signal-to-noise ratio.[223] Sequential ChIP assays (ChIP-reChIP)[224] and ChIP followed bisulfite sequencing (BSChIP-seq)[225,226] assays have been developed to identify the multiple binding events and determine whether these events are simultaneously present or occur on different chromosomes in the same cell or different cells. Numerous tools have been developed for ChIP-seq data analysis, and here, we review the computational processing pipelines emphasizing the essential steps of aligning the reads to the reference genome and detecting peaks.

Short-read alignment. During the ChIP procedure, the genomic DNA is sonicated or digested by MNase into a few hundred base pairs of DNA fragment, and during the sequencing procedure, 25–50 bp are sequenced at the two ends. Thus, short-read aligners must be fast and precise to locate their original position. There are two main strategies to achieve this goal: algorithms based on hash tables and algorithms based on suffix/prefix tries.[227] The classical BLAST,[228] ELAND (Illumina), SOAP,[229] MAQ,[230] RMAP,[163] and ZOOM[231] are hash table-based algorithms with different modifications on the spaced seed and sensitivity tolerance according to the reference genome. The algorithms based on suffix/prefix tries convert the inexact matching problem to the exact matching problem, which accelerate the computing speed remarkably. Of published aligners using this strategy, Bowtie,[232] BWA,[233] and SOAP2[234] are gaining increasing popularity. The choice of alignment method and the parameters selection such as mismatch allowance can impact the percentage of the successfully aligned reads, thus the next peak calling. More number of tools are summarized in Table 2.

Differential peak detection. The aligned unique ChIP-seq reads are usually identified as sets of enriched signals, termed peaks, on certain genomic regions. Data from DNA input control experiments are used as background levels of signal to compute the enrichment that would be expected by chance, thus pointing the position of the histone modification or protein binding sites. Peak detection requires a series of distinct steps before generating the final peak list as follows: reads shifting, background subtraction, peak identification, significance test, and artifacts removal.[235] Based on the signal characteristics, a variety of peak calling tools have been developed, and usually parameters in each step can be adjusted so that they dramatically affect the final peak. Histone modifications, histone variants, and histone-modifying enzymes usually give rise to diffuse signals and form peaks from several nucleosomes to large domains encompassing multiple genes, SICER[236] and BroadPeak[237] perform well under this circumstance. While for the exact binding locations of transcription factors and chromatin remodeling factors, MACS[238] and SISSRS[239] are of good achievements. There are comparison analyses for different peak callers, which may provide critical assessment idea when handling ChIP-seq data.[240,241] Comprehensive peak detection tools are listed in Table 2.

ncRNAs

ncRNA discovery and quantification. Transcriptome studies have confirmed that the genome sequences are greatly transcribed, and the vast amount of genetic information transcribed indicates that there are hidden categories, and the functions and biological significance of ncRNAs remain unclear. Rapidly accumulating evidences suggest that ncRNAs act as regulatory molecules in an epigenetic manner that associate with almost all biological processes,[242–245] and the complexity of the regulatory mechanisms stays in line with the complexity of organisms.[246] With the wide use of the NGS, it is becoming more powerful to discover new classes of ncRNA and investigate their functions from deep sequencing data. Earlier ncRNA endeavors were based on machine learning methods prediction and experimental validation, which were based on the ncRNA features such as evolutionary sequence conservation, RNA secondary structure and distinct expression patterns across developmental stages, different tissues, and conditions.[247] It has been proved that integrated analyzing of the RNA sequence, structure, and expression feature enables the ncRNA differentiation from protein-coding RNAs and regulatory elements and potentially different ncRNA categories,[248,249] so that paving a way to detect novel ncRNAs from unannotated genomic regions with systematic searching. Recently developed high-throughput ncRNA sequencing data analysis tools are emerging as systematic analysis pipelines, which are usually compromising three main aspects including ncRNA identification and quantification, interactions with RBPs and target mRNAs, and function characterization. The general workflows are first filtering the adapters and aligning the deep sequencing reads or conducting the de novo

assembly, next the known ncRNAs quantification and novel ncRNAs identification by inferring annotation databases, following the functional interaction analysis based on structural features and database annotations. iMir[250] is such an integrated pipeline with graphical user interface (GUI) that allows ncRNAs' identification such as miRNA and piRNA by miRAnalyzer[251] or miRDeep2,[252] differential expression analysis by DESeq,[253] and prediction of target using TargetScan[254] and miRanda.[255] Besides the alignments, quantification of known ncRNAs, CAP-miRSeq,[256] can detect and quantify precursor, mature, and novel miRNAs, analyze differential expressions by edgeR,[257] detect single-nucleotide variants (SNVs) by Genome Analysis Tool Kit (GATK),[258] which represents a unique feature of this kind of pipelines, and visualize by IGV genome browser. omiRas[259] and UEA sRNA workbench[260] can take the raw small ncRNA seq data and visualize the ncRNAs interaction network through a web service leveraging on several miRNA–mRNA databases after differential expression and comprehensive analysis.

LncRNA discovery and analysis has also been promoted by deep sequencing technology, while the challenges are the sensitivity and specificity of the detection due to the low expression level comparing to the protein-coding RNAs and limited annotation, so that it is difficult to uncover the biological functions. iSeeRNA[261] is a support vector machine (SVM)-based classifier that utilized lncRNA features of conservation, ORFs, and sequences characteristics to precisely separate them from coding genes. Self-estimation-based novel lincRNA filtering (Sebnif)[262] accurately detects lincRNAs through filtering the known and unknown, single-exon and multi-exon, size between 200 bp and 10 kb and other features based on iSeeRNA and annotates the detected lincRNAs with weighted gene coexpression network.[263] Based on the idea that similar expression patterns across different conditions may share similar functions and biological pathways, LncRNA2Function[264] provides an approach to annotate lncRNA by calculating the Pearson correlation coefficient (PCC) of lncRNA–mRNA pairs for the 10,000 lncRNAs in GENCODE project.[265]

ncRNA and protein interactions detection. Besides the chromatin modifications' regulation on gene expression, post-transcriptional mechanisms play a crucial role to tune the RNA level and protein level. A principal mechanism under intensive study is the RBP binding and action mechanism, for which CLIP-seq protocols enable the transcriptome-wide examination of interaction regions for particular RBPs. Hence, computational data analysis is key to the further understanding of transcriptome level regulation mechanisms. CLIP-seq generates the selected short reads from the RBPs binding regions, so that the reads alignment or reads mapping to the genome and transcriptome are usually the first step of the data analysis pipelines. Many software developed for genomic sequencing reads mapping can be directly implemented such as Bowtie, RMAP,[266] and Novoalign.[267] The mapping tools that consider the splicing and can detect the exon–exon junctions are also commonly used, which includes TopHat[268] and STAR.[269] It is worth to note that at least one nucleotide mismatch should be allowed in alignment especially for the PAR-CLIP sequencing data since the cross-link step can induce the T to C transition. After the reads mapping and cluster detection, the following step will be peak calling and binding site detection, which greatly depend on the transcript abundance and cluster length. The most commonly used strategy for this step is to find the precise cluster distribution profiles through enhancing the signal-to-noise ratio and decreasing the false-positive rate. Data analysis methods developed for this purpose include PIPE-CLIP,[270] PARalyzer,[271] Piranha,[266] wavClusteR,[272] and dCLIP.[273] The next downstream of the pipeline is the motif discovery, higher level structure prediction, and functional characterization. Previously developed tools for DNA and protein motif discoveries can be implemented to the RNA datasets and performed well, which include HOMER,[274] MEME,[275] cERMIT,[276] GLAM2,[277] MatrixREDUCE,[278] and RNA-context.[279] Although there are tools for the ncRNA secondary structure prediction and functional annotation, such as GraphProt,[280] CapR,[281] and LncRNA2Function,[264] there are still significant challenges in this field including increasing the sensitivity and specificity, decreasing the false-positive discovery rate, and expanding the algorithms for global prediction. After a complete understanding of the ncRNA and RBP regulatory mechanism is achieved, integrative approaches for a network-level interference can be explored.

Storing, Retrieving, and Visualizing Epigenomics Data

Once the most fundamental analysis of epigenomics data, including reads mapping and either DMR or peak calling, have been completed, the next main step is to store, retrieve, and visualize the epigenomic data across the sample groups. A common interest is to inspect or compare the DNA methylation and histone modification levels in a selected genomic region, such as gene locus, regulatory regions by a genome browser, either a Web-based genome browser (such as UCSC Genome Browser,[282] Ensembl,[283] WashU Human Epigenome Browser,[284] DaVIE[285]) or desktop-based local genome browser (such as IGV[286]). To do so, the specialized format files such as bigBed or bigWig converted from the BED or WIG files are required to be uploaded or imported into a genome browser. Among them, UCSC Genome Browser is widely used by allowing uploading custom tracks as well as displaying the tracks publicly. A general user can store the large volumes of epigenetic data in Gene Expression Omnibus[287] (GEO) from National Center for Biotechnology Information (NCBI) or DaVIE.[285] Additionally, several large-scale initiatives host the data in the public hub, as described in the next section. Researchers can retrieve the datasets from either these public hubs or the specialized databases, such as MethyBase,[183] MethDB,[288] MethyCancer,[289] and PubMeth[290]

Table 2. Software and tools for epigenomic data analysis.

SOFTWARE/TOOL	DESCRIPTION	URL	REFS
1. DNA methylation			
1.1. Mapping BS-seq reads			
1.1.1. General aligners with a BS-Seq module			
GSNAP	A wild-card bisulfite aligner included in a general-purpose alignment tool (Genomic Short-read Nucleotide Alignment Program)	http://share.gene.com/gmap	323
LAST	A wild-card bisulfite aligner included in a general-purpose alignment tool	http://last.cbrc.jp	161
RMAP	A Wild-card bisulfite aligner included in a general-purpose alignment tool	http://rulai.cshl.edu/rmap/	6
segemehl	A wild-card bisulfite aligner included in a general-purpose alignment tool	http://www.bioinf.uni-leipzig.de/Software/segemehl	304
1.1.2 Specific BS-Seq aligner that use a three-letter approach			
Bismark	A widely used three-letter bisulfite aligner based on Bowtie/Bowtie2	http://www.bioinformatics.babraham.ac.uk/projects/bismark	165
BRAT	A bisulfite-treated reads tool using the three-letter alignment	http://compbio.cs.ucr.edu/brat	166
BS-Seeker	A three-letter bisulfite aligner based on Bowtie	https://github.com/BSSeeker/Bsseeker2	324
MethylCoder	A three-letter bisulfite aligner based on Bowtie/GSNAP	https://github.com/brentp/methylcode	168
1.1.3 The specific BS-Seq aligner by wild-card approch			
BSMAP	A widely used wild-card aligner for bisulfite sequencing reads	http://code.google.com/p/bsmap	325
Pash	A wild-card bisulfite aligner using gapped k-mer and multi-positional hash table	http://brl.bcm.tmc.edu/pash	170–172
1.1.4 Other BS-seq aligners			
BISMA	Mapping and clustering of bisulfite sequencing data for individual clones from unique and repetitive sequences	http://biochem.jacobs-university.de/BDPC/BISMA/	326
BRAT-BW	A fast, accurate and memory-efficient BS aligner using the FM-index (Burrows-Wheeler transform)	http://compbio.cs.ucr.edu/brat/	304
B-SOLANA	A aligner for bisulfite-sequencing data of ABI SOLiD sequencers	http://code.google.com/p/bsolana	327
RRBSMAP	A wild-card aligner for RRBS reads	http://rrbsmap.computational-epigenetics.org	328
1.2. Detecting differential methylated regions (DMRs)			
1.2.1 Software for DMR calling only			
BiSeq	An R package for detect differentially methylated regions (DMRs) for BS data	https://www.bioconductor.org/packages/release/bioc/html/BiSeq.html	175
bumphunter	Bump hunting to identify differentially methylated regions	http://bioconductor.org/packages/release/bioc/html/bumphunter.html	177
DMRcate	An R package for detecting differentially methylated regions (DMRs) based on tunable kernel smoothing	www.bioconductor.org/packages/release/bioc/html/DMRcate.html	178
IMA	An R package for high-throughput analysis of Illumina's 450K Infinium methylation data	http://www.rforge.net/IMA	329
M3D	An R package for detecting differentially methylated regions (DMRs) using a non-parametric, kernel-based method	https://www.bioconductor.org/packages/release/bioc/html/M3D.html	330

(*Continued*)

Table 2. (*Continued*)

SOFTWARE/TOOL	DESCRIPTION	URL	REFS
methylSig	An R package for detecting differentially methylated sites (DMCs) or regions (DMRs) using a beta-binomial model	https://github.com/sartorlab/methylSig	331
metilene	A fast and sensitive tool for detecting DMR by a binary segmentation algorithm combined with a two-dimensional statistical test	http://www.bioinf.uni-leipzig.de/Software/metilene/	185
MOABS	A tool for detecting differentially methylated sites (DMCs) or regions (DMRs) based on a Beta-Binomial hierarchical model with relative low CpG coverage (~10X)	https://code.google.com/archive/p/moabs/	332
NHMMfdr	An R package for detecting differential DNA methylation based on non-homogeneeous hidden Markov model (NHMM) by estimating false discovery rates (FDRs)	http://www.ams.sunysb.edu/~pfkuan/NHMMfdr/	182
QDMR	A tool for detecting DMR based on Shannon entropy	http://bioinfo.hrbmu.edu.cn/qdmr	333
1.2.2 Pipeline for both BS-seq mapping and DMR calling			
Bsmooth	Bsmooth is a pipeline for analyzing whole genome bisulfite sequencing (WGBS) data. It includes tools for aligning the data, quality control, and identifying differentially methylated regions (DMRs).	http://rafalab.jhsph.edu/bsmooth/	304
MethPipe	A computational pipeline for analyzing bisulfite sequencing data (WGBS and RRBS), including BS mapping (Wild-Card aligner) and DMR calling	http://smithlabresearch.org/software/methpipe/	334
RefFreeDMA	Mapping for RRBS reads and DMR calling without a reference genome	https://github.com/jklughammer/RefFreeDMA	335
2. Histone Modifications and DNA-binding Proteins			
2.1 Short-read Alignment			
BWA	A fast and efficientlight-weighted tool that aligns short sequences to a sequence database; based on the Burrows–Wheeler transform	http://bio-bwa.sourceforge.net	233
Bowtie	Ultrafast, memory-efficient short read aligner. Uses a Burrows-Wheeler-Transformed (BWT) index	http://bowtie-bio.sourceforge.net	232
ELAND	Efficient Large-Scale Alignment of Nucleotide Databases. Whole genome alignments to a reference genome	http://support.illumina.com/help/SequencingAnalysisWorkflow/Content/Vault/Informatics/Sequencing_Analysis/CASAVA/swSEQ_mCA_ReferenceFiles.htm	Illumina
GenomeMapper	GenomeMapper is a short read mapping tool designed for accurate read alignments. It quickly aligns millions of reads either with ungapped or gapped alignments	http://1001genomes.org/software/genomemapper.html	336
GNUMAP	Genomic Next-generation Universal MAPper is a program designed to accurately map sequence data obtained from next-generation sequencing machines back to a genome of any size. It seeks to align reads from nonunique repeats using statistics	http://dna.cs.byu.edu/gnumap/	323
HiCUP	A tool for mapping and performing quality control on Hi-C data	http://www.bioinformatics.babraham.ac.uk/projects/hicup/	337
GSNAP	Considers a set of variant allele inputs to better align to heterozygous sites	http://research-pub.gene.com/gmap	160

(*Continued*)

Table 2. (*Continued*)

SOFTWARE/TOOL	DESCRIPTION	URL	REFS
MAQ	Mapping and Assembly with Qualities (renamed from MAPASS2). Particularly designed for Illumina with preliminary functions to handle ABI SOLiD data	http://maq.sourceforge.net/	230
SOAP	SOAP (Short Oligonucleotide Alignment Program). A program for efficient gapped and ungapped alignment of short oligonucleotides onto reference sequences	http://soap.genomics.org.cn/	229
SOAP2	SOAP2 used a Burrows Wheeler Transformation (BWT) compression index to substitute the seed strategy for indexing the reference sequence in the main memory	http://soap.genomics.org.cn/soapaligner.html	234
ZOOM	ZOOM (Zillions Of Oligos Mapped) is designed to map millions of short reads, emerged by next-generation sequencing technology, back to the reference genomes, and carry out post-analysis	http://omictools.com/zoom-tool	231
2.2 Peak Detection			
2.2.1 Peak Caller			
BroadPeak	A novel algorithm for identifying broad peaks in diffuse ChIP-seq datasets	http://jordan.biology.gatech.edu/page/software/broadpeak/	237
MACS	MACS fits data to a dynamic Poisson distribution; works with and without control data	http://liulab.dfci.harvard.edu/MACS	238
PeakSeq	PeakSeq takes into account differences in mappability of genomic regions; enrichment based on FDR calculation	http://info.gersteinlab.org/PeakSeq	338
SICER	A clustering approach for identification of enriched domains from histone modification ChIP-Seq data	http://home.gwu.edu/~wpeng/Software.htm	236
SISSRS	A novel algorithm for precise identification of binding sites from short reads generated from ChIP-Seq experiments	http://sissrs.rajajothi.com/	239
ZINBA	ZINBA can incorporate multiple genomic factors, such as mappability and GC content; can work with point-source and broad-source peak data	http://code.google.com/p/zinba	339
2.2.2 Differential Peak Caller			
baySeq	An R package that uses empirical Bayes approach to identify significant differences; assumes negative binomial distribution of data	http://www.bioconductor.org/packages/release/bioc/html/baySeq.html	340
ChIPDiff	A toolkit for the genome-wide comparison of histone modification sites identified by ChIP-seq, differential histone modification sites (DHMS) identification, uses binomial distribution, Baum-Welch expectation maximization (EM) algorithm, forward-backward algorithm	http://cmb.gis.a-star.edu.sg/ChIPSeq/paperChIP-Diff.htm	341
edgeR	An R package that uses negative binomial distribution to model differences in tag counts; uses replicates to better estimate significant differences	http://www.bioconductor.org/packages/2.9/bioc/html/edgeR.html	257

(*Continued*)

Table 2. (*Continued*)

SOFTWARE/TOOL	DESCRIPTION	URL	REFS
DESeq	DESeq uses negative binomial distribution, but differs in the calculation of the mean and variance of the distribution	http://www-huber.embl.de/users/anders/DESeq	253
SAMSeq	SAMSeq based on the popular SAM software; a non-parametric method that uses resampling to normalize for differences in sequencing depth	http://www.stanford.edu/~junli07/research.html#SAM	342
3. ncRNAs			
3.1 ncRNAs detection and quantification			
miRDeep	miRDeep was developed to discover active known or novel miRNAs from deep sequencing data after the removal of adapters with a number of scripts to preprocess and score the mapped data	https://www.mdc-berlin.de/8551903/en/	248
miRDeep2	miRDeep2 is more sensitively and robustly to carry out identifying known and novel miRNAs by evaluating the structure and signature for each precursor, quantifying known miRNAs based on the annotation in miRBase and predicting secondary structure by RNAfold tool	https://www.mdc-berlin.de/8551903/en/	252
miRDeep*	miRDeep* is an integrated standalone miRNA identification application with a user-friendly graphic interface to conduct sequence alignment, pre-miRNA secondary structure calculation, and graphical display with low memory requirement	http://www.australianprostatecentre.org/research/software/mirdeep-star	249
DARIO	DARIO is a web service for studying short read data from small RNA-seq experiments. It provides a wide range of analysis features, including quality control, read normalization, ncRNA quantification and prediction of putative ncRNA candidates	http://dario.bioinf.uni-leipzig.de/index.py	343
ncPRO-seq	ncPRO-seq is a tool for annotation and profiling of ncRNAs from small-RNA sequencing data. It aims to interrogate and perform detailed analysis on small RNAs derived from annotated non-coding regions in miRBase, piRBase, Rfam and repeatMasker, and regions defined by users. The ncPRO pipeline also has a module to identify regions significantly enriched with short reads that cannot be classified as known ncRNA families	https://sourceforge.net/projects/ncproseq/	344
CoRAL	CoRAL is a machine-learning package that can predict the precursor class of small RNAs present in a high-throughput RNA-sequencing dataset and produces information about the features that are most important for discriminating different populations of small non-coding RNAs	http://wanglab.pcbi.upenn.edu/coral/	345
RNA-CODE	RNA-CODE is designed for ncRNA identification in NGS data that lack quality reference genomes. Given a set of short reads, it classifies the reads into different types of ncRNA families. The classification results can be used to quantify the expression levels of different types of ncRNAs in RNA-seq data and ncRNA	http://www.cse.msu.edu/~chengy/RNA_CODE/	346

(*Continued*)

Table 2. (*Continued*)

SOFTWARE/TOOL	DESCRIPTION	URL	REFS
	composition profiles in metagenomic data, respectively		
CAP-miRSeq	A comprehensive analysis pipeline for deep microRNA sequencing that integrates read preprocessing, alignment, mature/precursor/novel miRNA qualification, variant detection in miRNA coding region, and flexible differential expression between experimental conditions	http://bioinformaticstools.mayo.edu/research/cap-mirseq/	256
iMir	A modular pipeline for comprehensive analysis of smallRNA-Seq data, comprising specific tools for adapter trimming, quality filtering, differential expression analysis, biological target prediction and other useful options by integrating multiple open source modules and resources in an automated workflow	http://www.labmedmolge.unisa.it/inglese/research/imir	250
UEA sRNA workbench	UEA sRNA workbench performs complete analysis of single or multiple-sample small RNA datasets to identify novel micro RNA sequences and profiling small RNA expression patterns in genetic data	http://srna-workbench.cmp.uea.ac.uk/	260
omiRas	omiRas is a web server for annotation, comparison and visualization of interaction networks of non-coding RNAs derived from small RNA-Sequencing	http://tools.genxpro.net/omiras/	259
sRNAtoolbox	sRNAtoolbox provide several tools including sRNAbench for sRNA expression profiling and prediction of novel microRNAs, sRNAde for differential expression analysis, miRNAconsTarget for prediction of miRNAs, sRNAjBrowserDE for visualization differential expression as a fuction of read length and sRNAfuncTerms for determination of over represented functional annotations in target gene set	http://bioinfo5.ugr.es/srnatoolbox	347
iSeeRNA	iSeeRNA is a support vector machine (SVM)-based classifier for the identification of lincRNAs	http://137.189.133.71/software.html	261
Sebnif	Sebnif is an Integrated Bioinformatics Pipeline for the Identification of Novel Large Intergenic Noncoding RNAs (lincRNAs) base on iSeeRNA	http://137.189.133.71/sebnif/	262
LncRNA2Function	LncRNA2Function – a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data	http://mlg.hit.edu.cn/lncrna2function/	264
3.2 RIP-seq and CLIP-seq			
3.2.1 Differential Peak Caller and Binding site detector from CLIP-seq			
Novoalign	An accurate NGS short reads aligner for aligning to reference genome	http://www.novocraft.com/products/novoalign/	267
PIPE-CLIP	A Galaxy framework-based comprehensive online pipeline for reliable analysis of data generated by three types of CLIP-seq protocol	http://pipeclip.qbrc.org/	270
PARalyzer	It utilizes this nucleotide ubstation in a kernel density estimate classifier to generate the high-resolution set of Protein-RNA interaction sites	https://ohlerlab.mdc-berlin.de/software/PARalyzer_85/	271

(*Continued*)

Table 2. (*Continued*)

SOFTWARE/TOOL	DESCRIPTION	URL	REFS
Piranha	Piranha is a peak finding and differential binding detection algorithm	http://smithlabresearch.org/software/piranha/	266
wavClusteR	An integrated pipeline for the analysis of PAR-CLIP data	https://bioconductor.org/packages/release/bioc/html/wavClusteR.html	272
dCLIP	dCLIP is designed for quantitative CLIP-seq comparative analysis is able to effectively identify differential binding regions of RBPs in four CLIP-seq datasets	http://qbrc.swmed.edu/software/	273
3.2.2 Motif Discovery			
GraphProt	GraphProt is a machine learning computational framework for learning sequence- and structure-binding preferences of RNA-RBPs from high-throughput experimental data	http://www.bioinf.uni-freiburg.de/Software/GraphProt/	280
MEME	Perform motif discovery on DNA, RNA or protein datasets	http://meme-suite.org/	348
cERMIT	cERMIT is a computationally efficient motif discovery tool based on analyzing genome-wide quantitative regulatory evidence	https://ohlerlab.mdc-berlin.de/software/cERMIT_82/	276
GLAM2 (Gapped Local Alignment of Motifs)	GLAM2 is a motif detection tool for discovering motifs allowing indels in a fully general manner from DNA, RNA and protein datasets	http://bioinformatics.org.au/glam2	277
MatrixREDUCE	A motif discovery tool for genome-wide ChIP-seq and CLIP-seq data analysis	http://www.bussemakerlab.org/	278
RNA Bind-n-Seq	A quantitative assessment of the sequence and structural binding specificity		349
CapR	An efficient algorithm that calculates the probability that each RNA base position is located within each secondary structural context	https://sites.google.com/site/fukunagatsu/software/capr	281
RNAcontext	An efficient motif finding method ideally suited for using large-scale RNA-binding affinity datasets to determine the relative binding preferences of RBPs for a wide range of RNA sequences and structures	http://www.cs.toronto.edu/~hilal/rnacontext/	279
ViennaRNA Package 2.0	A widely used compilation of RNA secondary structure	http://www.tbi.univie.ac.at/RNA/	279
4. Storing, retrieving and visualizing epigenomics data			
4.1 Genome browser for visualizing DNA methylation			
Ensembl	A widely used Web-based genome browser with various epigenome data sets	http://www.ensembl.org	283
IGV	A widely used graphical genome browser that is run locally on the user's computer	http://www.broadinstitute.org/igv	286
UCSC Genome Browser	Widely used Web-based genome browser hosting all ENCODE data	http://genome.ucsc.edu	282
BDPC	Web-based tool for bisulfite sequencing data presentation and compilation	http://biochem.jacobs-university.de/BDPC	350
DaVIE	The database with an intuitive user interface to perform visual comparisons across large DNA methylation data sets	https://github.com/apfejes/epigenetics-software	285

(*Continued*)

Table 2. (*Continued*)

SOFTWARE/TOOL	DESCRIPTION	URL	REFS
EpiExplorer	A web server provides an interactive gateway for exploring large-scale epigenetic datasets of the human and mouse genome	http://epiexplorer.mpi-inf.mpg.de	351
EpiGRAPH	A user-friendly software for advanced (epi-) genome analysis and prediction by powerful machine learning algorithms	http://epigraph.mpi-inf.mpg.de	352
WashU Epigenome Browser	Web-based genome browser focusing on the human epigenome	http://epigenomegateway.wustl.edu	353
4.2 Specialized-DNA methylation databases			
MethBase	A central reference methylome database created from public BS-seq datasets	http://smithlabresearch.org/software/methbase/	334
MethDB	A database for DNA methylation and environmental epigenetic effects	http://www.methdb.de	288
MethyCancer	Database of cancer DNA methylation data	http://methycancer.psych.ac.cn	354
PubMeth	Database of DNA methylation literature	http://www.pubmeth.org	290
4.3 Specialized histone modification databases			
ChromatinDB	A database of genome-wide histone modification patterns for *Saccharomyces cerevisiae*	http://integbio.jp/dbcatalog/en/record/nbdc00939?jtpl=56	294
CR Cistrome	A ChIP-Seq database for chromatin regulators and histone modification linkages in human and mouse	http://cistrome.org/cr/	293
Histome	A relational knowledgebase of human histone proteins and histone modifying enzymes	http://www.actrec.gov.in/histome/	292
HHMD	The human histone modification database	http://202.97.205.78/hhmd/	291
4.4 Specialized nc RNA and RBPs interaction database			
starBase V2.0	starBase is designed for decoding ncRNA and the RNA-protein interaction networks and predicting functions especially in cancer samples	http://starbase.sysu.edu.cn/	296,297
CLIPZ	CLIPZ supports the automatic functional annotation and visualization of CLIP-seq identified binding sites	http://www.clipz.unibas.ch/	298
doRiNA	A database of RNA interactions in post-transcriptional regulation	http://dorina.mdc-berlin.de/	300
CLIPdb	An intergrated resource for characterizing the regulatory networks between RBPs and various RNA transcript classes	http://lulab.life.tsinghua.edu.cn/clipdb/	301

Note: *The descriptions are adapted from the software/tools website descriptions.

for DNA methylation data and HHMD,[291] Histome,[292] CR Cistrome,[293] and ChromatinDB[294] for histone modification data. Accumulating CLIP-seq data generation and analysis improves the annotation in terms of ncRNA category, secondary structure, and RBP binding regions. Besides the GEO and ArrayExpress form European Bioinformatics Institute (EBI),[295] there are currently four databases focusing on ncRNA and RBPs binding information as follows: starBase V2.0,[296,297] CLIPZ,[298] doRiNA,[299,300] and CLIPdb.[301] These epigenomic data storing and visualizing websites, tools, and databases are summarized in Table 2.

Public Resources for Reference Epigenome

Data analysis and algorithm development have been accelerated by several collaborative projects in addition to data generation, which has leveraged by experimental pipelines built around NGS technologies. Multiple comprehensive epigenomic projects have been launched with the goal of providing

a public resource and delivering a collection of normal epigenomes as a reference framework to catalyze basic biology and disease-oriented research.

The International Human Epigenome Consortium (IHEC)[302] (http://ihec-epigenomes.org/) launched with "a goal to understand to what extent the epigenome has shaped the human population genetically and in response to the environment by coordinating the reference maps of human epigenomes for key cellular states in health and diseases status". It has been distributed to multiple contributing projects including the NIH roadmap, the Encyclopedia of DNA Elements (ENCODE), and the BLUEPRINT projects. The NIH Roadmap Epigenomics Mapping Consortium[303] (http://www.roadmapepigenomics.org/) aims to deliver a collection of normal epigenomes to provide a framework or reference for comparison and integration within a broad array of future studies. The Consortium has mapped DNA methylation, histone modifications, chromatin accessibility, and small RNA transcripts in representative tissues and cell lines that are the normal counterparts of tissues and organ systems frequently involved in human disease. The ENCODE[304] (https://www.encodeproject.org/) and the modENCODE (http://www.modencode.org/) projects are dedicated to list all functional elements for gene expression regulation in the genome of human and model organisms by integrating epigenomic, transcriptomic, genomic, and proteomics data. It provides extensive epigenome data for cultured cell lines in addition if IHEC focus on primary cell types. GENCODE project (https://www.gencodegenes.org)[305] is a scale-up of the ENCODE project for integrated annotation of gene features in human and mouse. The endeavor focuses on accurate annotation with all evidence-based gene features including protein-coding loci with alternatively spliced variants, noncoding loci, and pseudogenes. The European BLUEPRINT project (http://www.blueprint-epigenome.eu/) studies a variety of blood cell types and their associated diseases, and the German DEEP project (http://www.deutsches-epigenom-programm.de/) analyzes cell types that are related to metabolic and inflammatory diseases with high socioeconomic impact. The Human Epigenome Project (HEP) (http://www.epigenome.org/) focuses on the genome-wide DNA methylation pattern identification, catalog, and interpretation in all human genes with deciphering methylation variable positions (MVPs) to promise significant advance of human disease understanding and diagnoses. All IHEC data are being distributed via its GEO database in the global bioinformatics hubs of the US NCBI, and its European Genome–Phenome Archive (EGA) database in the EBI. FANTOM project (http://fantom.org/)[131] is the first large-scale catalog for ncRNAs, from which over 67,000 cDNAs have been sequenced and 3,652 with confident experimental evidences.[306]

Meanwhile, cancer genomic projects including The Cancer Genome Atlas (TCGA)[307] (http://cancergenome.nih.gov/) and International Cancer Genome Consortium (ICGC)[308] (https://icgc.org/) aim to obtain a comprehensive and multidimensional description of genomic, transcriptomic, and epigenomic changes in different tumor types and help understanding what errors cause cells grow uncontrolled, how the cancer can be prevented, early diagnosis, and better treatment. Aggregated data are freely accessible from the TCGA Data Portal and ICGC Data Portal, but an application is required to access raw sequencing data and genotype information of individual patients. More comprehensive projects initiated in institutional and regional-wide that provide epigenomic data resources are listed in Table 3.

Integration Analysis

The experimental and bioinformatics methods for epigenetics data analysis have undergone a revolution in the past decade along with the advances of NGS technology, especially in the throughput and multiple dimension of detection. Over the coming years, with more epigenomics data becoming available through public consortia, researchers can investigate the biological process and disease in a comprehensive way by mapping the DNA methylation, histone modifications, transcription factor binding, nucleosome positioning, and chromosomal organization combined with transcriptomic and proteomic data.[309–311] Simple integration analysis is intersection analysis among features extracted from different approaches, such as histone modification data from ChIP-seq, DNA methylation data from BS-seq, and gene expression data from RNA-seq, exome, or whole-genome sequencing,[128] which may facilitate understanding of developmental event or disease study. In addition, combination of different datasets from multiple projects and study of the feature of more subjects increasingly requires the integrative analysis.

Biological systems are being deeply investigated at an unprecedented scale along with the rise of novel omics technologies and through large-scale consortia projects. However, the heterogeneity and large volume of these datasets are still obstacles of the integrative analysis, which encourage researchers to develop novel data integration methodologies.

Outlook

The advances of epigenomic study with NGS development has profoundly challenged the long-held traditional view of the genetic code being the key determinant of gene function and its alteration being the major cause of complex diseases. Advances in the epigenetic field have led to the realization that the packaging of the genome is as important as the genome sequence in regulating fundamental cellular processes and its alteration being an essential cause of human diseases. Comprehensive understandings of the global patterns and the interplays of these epigenetic regulators and their corresponding changes upon environmental stimuli have enabled the better understanding of biology and better diagnosis and treatment strategies for diseases.

Computational analysis in epigenomics holding the great power of helping in revealing genome-wide landscape and

Table 3. Large-scale epigenome projects.

PROJECTS AND WEBSITES	SUMMARY
IHEC (International Human Epigenome Consortium) (http://ihec-epigenomes.org)	IHEC launched with a goal to understand to what extent the epigenome has shaped the human population genetically and in response to the environment by coordinating the reference maps of human epigenomes for key cellular states in health and diseases status. It has been distributed to multiple contributing projects including the NIH Roadmap, the ENCODE and the BLUEPRINT projects.
NIH Roadmap Epigenomics Mapping Consortium (http://www.roadmapepigenomics.org/)	The NIH Roadmap Epigenomics Mapping Consortium was launched with the goal of producing a public resource of human epigenomic data to catalyze basic biology and disease-oriented research. The Consortium expects to deliver a collection of normal epigenomes that will provide a framework or reference for comparison and integration within a broad array of future studies.
ENCODE (Encyclopedia of DNA Elements) (https://www.encodeproject.org/)	The ENCODE Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active. Although epigenome mapping is not its main goal, the project includes largescale mapping of DNA methylation, histone modifications and other epigenetic information.
BLUEPRINT (http://www.blueprint-epigenome.eu/)	BLUEPRINT is a large-scale research project receiving close to 30 million euro funding from the EU. 39 leading European universities, research institutes and industry entrepreneurs participate in what is one of the two first so-called high impact research initiatives to receive funding from the EU.
HEP (Human Epigenome Project) (http://www.epigenome.org/)	The partially EU-funded HEP analyzed DNA methylation in 43 unrelated individuals at single basepair resolution. Although the analysis was confined to selected regions on three chromosomes, it is the largest high-resolution, multiindividual epigenome dataset published to date.
German DEEP project (http://www.deutsches-epigenom-programm.de/)	DEEP focuses on the analysis of cells connected to complex diseases with high socio-economic impact: metabolic diseases such as steatosis and adipositas as well as inflammatory diseases of the joints and the intestine. DEEPs goal is to generate high-end data for comprehensive biomedical interpretation of healthy and diseased cells. With this DEEP will contribute to discover new functional epigenetic links useful for clinical diagnosis, therapy and health risk prevention. All data generated will be made publically available and will be integrated into a sustainable worldwide data structure comprised by the IHEC initiative.
HEROIC (High-throughput Epigenetic Regulatory Organisation In Chromatin) (EU) (http://projects.ensembl.org/heroic/)	The HEROIC project is a multi-center EU project that applies ChIP-on-chip, chromosome interaction analysis and whole-genome nuclear localization assays to understanding human genome regulation.
AHEAD (Alliance for Human Epigenomics and Disease) Task Force (international) (http://graphy21.blogspot.com)	The goal of the AHEAD is to initiate and coordinate a comprehensive human epigenome-mapping project. Initially, focus is set on developing a suitable bioinformatic infrastructure and on performing epigenome mapping in a selection of normal tissues, which may provide the reference for subsequent mapping in abnormal cells.
ICGC (International Cancer Genome Consortium) (https://icgc.org/)	The goal of the ICGC is to obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.
TCGA (The Cancer Genome Atlas) (http://cancergenome.nih.gov)	The Cancer Genome Atlas (TCGA), collaboration between the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI), aims to generate comprehensive, multi-dimensional maps of the key genomic changes in major types and subtypes of cancer.
FANTOM project (http://fantom.gsc.riken.jp/)	FANTOM is an international research consortium established to assign functional annotations to the full-length cDNAs that were collected during the Mouse Encyclopedia Project at RIKEN. FANTOM developed and expanded over time to encompass the fields of transcriptome analysis. FANTOM database and the FANTOM full-length cDNA clone bank are worldwide available resources that already fueled the iPS development.
GENECODE project (https://www.gencodegenes.org/)	GENCODE as a sub-project of the ENCODE scale-up project are aiming to integrated annotation of gene features. Currently running phase is continuously to improve the coverage and accuracy of the human and mouse gene set by enhancing and extending the annotation of all evidence-based gene features at a high accuracy, including protein-coding loci with alternatively splices variants, non-coding loci and pseudogenes.

Note: *The descriptions are adapted from indicated website sources.

interplay with genome will significantly increase in the coming years. It will increasingly take both the genome sequence and the proteins interacted with the genome into account as regulatory networks for the cellular processes. The decreasing cost of the NGS technologies will enable quantitative analysis of epigenetic variation from single-cell level to human individuals in a population level, which will greatly facilitate precision medicine and analysis of various effects of environmental factors on the human genome. Computational epigenomics data analysis will also promote the development of theoretical models and powerful tools, which will in turn facilitate further investigations toward the depiction of big picture of life science.

Acknowledgments

We thank the NIH Fellows Editorial Board for editorial suggestions on the manuscript. The content of this publication neither necessarily reflects the views or policies of the Department of Health and Human Services nor mentions trade names, commercial products, or organizations that imply endorsement by the US Government.

Author Contributions

Wrote the first draft of the manuscript: YH, XH. Contributed to the writing of the manuscript: YH, XH. Agree with manuscript results and conclusions: YH, XH. Jointly developed the structure and arguments for the paper: YH. Made critical revisions and approved final version: YH, XH. Both authors reviewed and approved of the final manuscript.

REFERENCES

1. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. *Genes Dev*. 2009;23(7):781–3.
2. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*. 2007;447(7143):425–32.
3. Richards EJ. Inherited epigenetic variation – revisiting soft inheritance. *Nat Rev Genet*. 2006;7(5):395–401.
4. Feinberg AP, Irizarry RA. Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci USA*. 2010;107(suppl 1):1757–64.
5. 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):170–9.
6. Kong A, Steinthorsdottir V, Masson G, et al. Parental origin of sequence variants associated with complex diseases. *Nature*. 2009;462(7275):868–74.
7. Brookes E, Shi Y. Diverse epigenetic mechanisms of human disease. *Annu Rev Genet*. 2014;48:237–68.
8. Golbabapour S, Abdulla MA, Hajrezaei M. A concise review on epigenetic regulation: insight into molecular mechanisms. *Int J Mol Sci*. 2011;12(12):8661–94.
9. Robinson MD, Pelizzola M. Computational epigenomics: challenges and opportunities. *Front Genet*. 2015;6:88.
10. Solter D. Imprinting today: end of the beginning or beginning of the end? *Cytogenet Genome Res*. 2006;113(1–4):12–6.
11. Ushijima T, Watanabe N, Okochi E, Kaneda A, Sugimura T, Miyamoto K. Fidelity of the methylation pattern and its variation in the genome. *Genome Res*. 2003;13(5):868–74.
12. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics*. 1998;148(4):1667–86.
13. Woodcock CL. Chromatin architecture. *Curr Opin Struct Biol*. 2006;16(2):213–20.
14. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev*. 2002;16(1):6–21.
15. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128(4):693–705.
16. Ren J, Briones V, Barbour S, et al. The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. *Nucleic Acids Res*. 2015;43(3):1444–55.
17. Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*. 2013;154(3):490–503.
18. Shen X, Xiao H, Ranallo R, Wu WH, Wu C. Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. *Science*. 2003;299(5603):112–4.
19. Vignali M, Hassan AH, Neely KE, Workman JL. ATP-dependent chromatin-remodeling complexes. *Mol Cell Biol*. 2000;20(6):1899–910.
20. Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov*. 2012;11(5):384–400.
21. Falkenberg KJ, Johnstone RW. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat Rev Drug Discov*. 2014;13(9):673–91.
22. Musselman CA, Lalonde ME, Cote J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol*. 2012;19(12):1218–27.
23. Morris KV, Mattick JS. The rise of regulatory RNA. *Nat Rev Genet*. 2014;15(6):423–37.
24. Qureshi IA, Mehler MF. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nat Rev Neurosci*. 2012;13(8):528–41.
25. Consortium EP, Birney E, Stamatoyannopoulos JA, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799–816.
26. Bertone P, Stolc V, Royce TE, et al. Global identification of human transcribed sequences with genome tiling arrays. *Science*. 2004;306(5705):2242–6.
27. Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*. 1996;85(7):1077–88.
28. Ni J, Tien AL, Fournier MJ. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell*. 1997;89(4):565–73.
29. Seila AC, Calabrese JM, Levine SS, et al. Divergent transcription from active promoters. *Science*. 2008;322(5909):1849–51.
30. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116(2):281–97.
31. Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. *Cell*. 2009;136(4):642–55.
32. Malone CD, Brennecke J, Dus M, et al. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell*. 2009;137(3):522–35.
33. Li Q, Hu B, Hu GW, et al. tRNA-Derived small non-coding RNAs in response to ischemia inhibit angiogenesis. *Sci Rep*. 2016;6:20850.
34. Selitsky SR, Baran-Gale J, Honda M, et al. Small tRNA-derived RNAs are increased and more abundant than microRNAs in chronic hepatitis B and C. *Sci Rep*. 2015;5:7675.
35. Lim LP, Lau NC, Garrett-Engele P, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*. 2005;433(7027):769–73.
36. Golden DE, Gerbasi VR, Sontheimer EJ. An inside job for siRNAs. *Mol Cell*. 2008;31(3):309–12.
37. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science*. 2008;322(5906):1387–92.
38. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*. 2014;157(1):77–94.
39. Kondo T, Plaza S, Zanet J, et al. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science*. 2010;329(5989):336–9.
40. Hung T, Wang Y, Lin MF, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011;43(7):621–9.
41. Yin QF, Yang L, Zhang Y, et al. Long noncoding RNAs with snoRNA ends. *Mol Cell*. 2012;48(2):219–30.
42. Orom UA, Derrien T, Beringer M, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell*. 2010;143(1):46–58.
43. Bertani S, Sauer S, Bolotin E, Sauer F. The noncoding RNA mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol Cell*. 2011;43(6):1040–6.
44. Wang KC, Yang YW, Liu B, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011;472(7341):120–4.
45. Lai F, Orom UA, Cesaroni M, et al. Activating RNAs associate with mediator to enhance chromatin architecture and transcription. *Nature*. 2013;494(7438):497–501.
46. Nelson BR, Makarewich CA, Anderson DM, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*. 2016;351(6270):271–5.

47. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. *Cell*. 2009;136(4):656–68.
48. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013;154(1):26–46.
49. Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev*. 2009;23(16):1831–42.
50. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet*. 2013;14(3):204–20.
51. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*. 2008;9(6):465–76.
52. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*. 2011;12(1):7–18.
53. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet*. 2016;17(9):551–65.
54. Bickmore WA, van Steensel B. Genome architecture: domain organization of interphase chromosomes. *Cell*. 2013;152(6):1270–84.
55. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485(7398):376–80.
56. Cheutin T, Cavalli G. Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr Opin Genet Dev*. 2014;25:30–7.
57. Dixon JR, Jung I, Selvaraj S, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518(7539):331–6.
58. Jager R, Migliorini G, Henrion M, et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat Commun*. 2015;6:6178.
59. Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol*. 2010;28(10):1057–68.
60. Gupta B, Hawkins RD. Epigenomics of autoimmune diseases. *Immunol Cell Biol*. 2015;93(3):271–6.
61. Salpea P, Russanova VR, Hirai TH, et al. Postnatal development- and age-related changes in DNA-methylation patterns in the human genome. *Nucleic Acids Res*. 2012;40(14):6477–94.
62. Han Y, Han D, Yan Z, et al. Stress-associated H3 K4 methylation accumulates during postnatal development and aging of rhesus macaque brain. *Aging Cell*. 2012;11(6):1055–64.
63. Putiri EL, Robertson KD. Epigenetic mechanisms and genome stability. *Clin Epigenetics*. 2011;2(2):299–314.
64. Benayoun BA, Pollina EA, Brunet A. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat Rev Mol Cell Biol*. 2015;16(10):593–610.
65. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA*. 2005;102(30):10604–9.
66. Barbot W, Dupressoir A, Lazar V, Heidmann T. Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic Acids Res*. 2002;30(11):2365–73.
67. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*. 2007;8(4):286–98.
68. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP. Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science*. 1985;228(4696):187–90.
69. Gaudet F, Hodgson JG, Eden A, et al. Induction of tumors in mice by genomic hypomethylation. *Science*. 2003;300(5618):489–92.
70. Miremadi A, Oestergaard MZ, Pharoah PD, Caldas C. Cancer genetics of epigenetic genes. *Hum Mol Genet*. 2007;16(Spec No 1):R28–49.
71. Lin RK, Wang YC. Dysregulated transcriptional and post-translational control of DNA methyltransferases in cancer. *Cell Biosci*. 2014;4:46.
72. Casillas MA Jr, Lopatina N, Andrews LG, Tollefsbol TO. Transcriptional control of the DNA methyltransferases is altered in aging and neoplastically-transformed human fibroblasts. *Mol Cell Biochem*. 2003;252(1–2):33–43.
73. Warton K, Samimi G. Methylation of cell-free circulating DNA in the diagnosis of cancer. *Front Mol Biosci*. 2015;2:13.
74. Delpu Y, Cordelier P, Cho WC, Torrisani J. DNA methylation and cancer diagnosis. *Int J Mol Sci*. 2013;14(7):15029–58.
75. Lorincz AT. Cancer diagnostic classifiers based on quantitative DNA methylation. *Expert Rev Mol Diagn*. 2014;14(3):293–305.
76. Si X, Zhao Y, Yang C, Zhang S, Zhang X. DNA methylation as a potential diagnosis indicator for rapid discrimination of rare cancer cells and normal cells. *Sci Rep*. 2015;5:11882.
77. Han S, Brunet A. Histone methylation makes its mark on longevity. *Trends Cell Biol*. 2012;22(1):42–9.
78. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis*. 2010;31(1):27–36.
79. Fraga MF, Ballestar E, Villar-Garea A, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet*. 2005;37(4):391–400.
80. Hamamoto R, Furukawa Y, Morita M, et al. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol*. 2004;6(8):731–40.
81. Kondo Y, Shen L, Suzuki S, et al. Alterations of DNA methylation and histone modifications contribute to gene silencing in hepatocellular carcinomas. *Hepatol Res*. 2007;37(11):974–83.
82. Vire E, Brenner C, Deplus R, et al. The polycomb group protein EZH2 directly controls DNA methylation. *Nature*. 2006;439(7078):871–4.
83. Dang W, Steffen KK, Perry R, et al. Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature*. 2009;459(7248):802–7.
84. 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):8824–9.
85. Sarg B, Koutzamani E, Helliger W, Rundquist I, Lindner HH. Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging. *J Biol Chem*. 2002;277(42):39195–201.
86. Vaquero A, Sternglanz R, Reinberg D. NAD+-dependent deacetylation of H4 lysine 16 by class III HDACs. *Oncogene*. 2007;26(37):5505–20.
87. Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*. 2010;463(7279):360–3.
88. Varambally S, Cao Q, Mani RS, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. *Science*. 2008;322(5908):1695–9.
89. Howitz KT, Bitterman KJ, Cohen HY, et al. Small molecule activators of sirtuins extend *Saccharomyces cerevisiae* lifespan. *Nature*. 2003;425(6954):191–6.
90. Wood JG, Rogina B, Lavu S, et al. Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature*. 2004;430(7000):686–9.
91. Kim D, Nguyen MD, Dobbin MM, et al. SIRT1 deacetylase protects against neurodegeneration in models for Alzheimer's disease and amyotrophic lateral sclerosis. *EMBO J*. 2007;26(13):3169–79.
92. Milne JC, Lambert PD, Schenk S, et al. Small molecule activators of SIRT1 as therapeutics for the treatment of type 2 diabetes. *Nature*. 2007;450(7170):712–6.
93. Zhang Q, Zeng SX, Zhang Y, et al. A small molecule inauhzin inhibits SIRT1 activity and suppresses tumour growth through activation of p53. *EMBO Mol Med*. 2012;4(4):298–312.
94. Hubbard BP, Sinclair DA. Small molecule SIRT1 activators for the treatment of aging and age-related diseases. *Trends Pharmacol Sci*. 2014;35(3):146–54.
95. Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet*. 2009;10(10):704–14.
96. Nicoloso MS, Spizzo R, Shimizu M, Rossi S, Calin GA. MicroRNAs – the micro steering wheel of tumour metastases. *Nat Rev Cancer*. 2009;9(4):293–302.
97. Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2002;99(24):15524–9.
98. Davalos V, Moutinho C, Villanueva A, et al. Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. *Oncogene*. 2012;31(16):2062–74.
99. Melo SA, Ropero S, Moutinho C, et al. A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat Genet*. 2009;41(3):365–70.
100. Hill DA, Ivanovich J, Priest JR, et al. DICER1 mutations in familial pleuropulmonary blastoma. *Science*. 2009;325(5943):965.
101. Hebert SS, Horre K, Nicolai L, et al. Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. *Proc Natl Acad Sci U S A*. 2008;105(17):6415–20.
102. Wang WX, Rajeev BW, Stromberg AJ, et al. The expression of microRNA miR-107 decreases early in Alzheimer's disease and may accelerate disease progression through regulation of beta-site amyloid precursor protein-cleaving enzyme 1. *J Neurosci*. 2008;28(5):1213–23.
103. Boissonneault V, Plante I, Rivest S, Provost P. MicroRNA-298 and microRNA-328 regulate expression of mouse beta-amyloid precursor protein-converting enzyme 1. *J Biol Chem*. 2009;284(4):1971–81.
104. Edbauer D, Neilson JR, Foster KA, et al. Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. *Neuron*. 2010;65(3):373–84.
105. Gehrke S, Imai Y, Sokol N, Lu B. Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. *Nature*. 2010;466(7306):637–41.
106. Glazer RI, Vo DT, Penalva LO. Musashi1: an RBP with versatile functions in normal and cancer stem cells. *Front Biosci (Landmark Ed)*. 2012;17:54–64.
107. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464(7291):1071–6.
108. Huarte M, Guttman M, Feldser D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010;142(3):409–19.
109. Eggermann T. Silver-Russell and Beckwith-Wiedemann syndromes: opposite (epi)mutations in 11p15 result in opposite clinical pictures. *Horm Res*. 2009;71(suppl 2):30–5.
110. Irizarry RA, Ladd-Acosta C, Carvalho B, et al. Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res*. 2008;18:780–90.

111. Estecio MR, Yan PS, Ibrahim AE, et al. High-throughput methylation profiling by MCA coupled to CpG island microarray. *Genome Res*. 2007;17:1529–36.
112. Jacinto FV, Ballestar E, Esteller M. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *Biotechniques*. 2008;44:35–43.
113. Rauch TA, Pfeifer GP. The MIRA method for DNA methylation analysis. *Methods Mol Biol*. 2009;507:65–75.
114. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet*. 2010;11:191–203.
115. Bock C. Analysing and interpreting DNA methylation data. *Nat Rev Genet*. 2012;13(10):705–19.
116. Lister R, Pelizzola M, Dowen RH, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009;462(7271):315–22.
117. Meissner A, Mikkelsen TS, Gu H, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*. 2008;454(7205):766–70.
118. Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129(4):823–37.
119. Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007;448(7153):553–60.
120. Pillai S, Chellappan SP. ChIP on chip assays: genome-wide analysis of transcription factor binding and histone modifications. *Methods Mol Biol*. 2009;523:341–66.
121. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007;316(5830):1497–502.
122. Mardis ER. ChIP-seq: welcome to the new frontier. *Nat Methods*. 2007;4(8):613–4.
123. 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):dbrot5384.
124. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002;295(5558):1306–11.
125. Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289–93.
126. Nagano T, Lubling Y, Stevens TJ, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 2013;502(7469):59–64.
127. Nagano T, Lubling Y, Yaffe E, et al. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat Protoc*. 2015;10(12):1986–2003.
128. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced applications of RNA sequencing and challenges. *Bioinform Biol Insights*. 2015;9(suppl 1):29–46.
129. Baran-Gale J, Kurtz CL, Erdos MR, et al. Addressing bias in small RNA library preparation for sequencing: a new protocol recovers microRNAs that evade capture by current methods. *Front Genet*. 2015;6:352.
130. Bono H, Kasukawa T, Furuno M, Hayashizaki Y, Okazaki Y. FANTOM DB: database of functional annotation of RIKEN mouse cDNA clones. *Nucleic Acids Res*. 2002;30(1):116–8.
131. Carninci P, Kasukawa T, Katayama S, et al. The transcriptional landscape of the mammalian genome. *Science*. 2005;309(5740):1559–63.
132. Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22(9):1775–89.
133. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res*. 2012;40(Database issue):D130-5.
134. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
135. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.
136. Streit S, Michalski CW, Erkan M, Kleeff J, Friess H. Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. *Nat Protoc*. 2009;4(1):37–43.
137. Chureau C, Chantalat S, Romito A, et al. Ftx is a non-coding RNA which affects xist expression and chromatin structure within the X-inactivation center region. *Hum Mol Genet*. 2011;20(4):705–18.
138. Siomi H, Siomi MC. On the road to reading the RNA-interference code. *Nature*. 2009;457(7228):396–404.
139. Tsai MC, Manor O, Wan Y, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010;329(5992):689–93.
140. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*. 2008;322(5902):750–6.
141. Jain R, Devine T, George AD, et al. RIP-chip analysis: RNA-binding protein immunoprecipitation-microarray (chip) profiling. *Methods Mol Biol*. 2011;703:247–63.
142. Zhao J, Ohsumi TK, Kung JT, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell*. 2010;40(6):939–53.
143. Licatalosi DD, Mele A, Fak JJ, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. 2008;456(7221):464–9.
144. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. 2009;460(7254):479–86.
145. Kaneko S, Bonasio R, Saldana-Meyer R, et al. Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell*. 2014;53(2):290–300.
146. Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495(7441):333–8.
147. Zhang X, Zuo X, Yang B, et al. MicroRNA directly enhances mitochondrial translation during muscle differentiation. *Cell*. 2014;158(3):607–19.
148. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*. 2011;44(4):667–78.
149. Lister R, Mukamel EA, Nery JR, et al. Global epigenomic reconfiguration during mammalian brain development. *Science*. 2013;341(6146):1237905.
150. Xie W, Barr CL, Kim A, et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell*. 2012;148(4):816–31.
151. Lister R, O'Malley RC, Tonti-Filippini J, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*. 2008;133(3):523–36.
152. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature*. 1986;321(6067):209–13.
153. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *J Mol Biol*. 1987;196(2):261–82.
154. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A*. 2002;99(6):3740–5.
155. Schmidl C, Klug M, Boeld TJ, et al. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res*. 2009;19(7):1165–74.
156. Rishi V, Bhattacharya P, Chatterjee R, et al. CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. *Proc Natl Acad Sci USA*. 2010;107(47):20311–6.
157. Eckhardt F. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*. 2006;38:1378–85.
158. Weber M, Hellmann I, Stadler MB, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 2007;39(4):457–66.
159. Kurdyukov S, Bullock M. DNA methylation analysis: choosing the right method. *Biology (Basel)*. 2016;5(1):E3.
160. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*. 2010;26:873–81.
161. Frith MC, Mori R, Asai K. A mostly traditional approach improves alignment of bisulfite-converted DNA. *Nucleic Acids Res*. 2012;40:e100.
162. Mann IK, Chatterjee R, Zhao J, et al. CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome Res*. 2013;23(6):988–97.
163. Smith AD, Chung WY, Hodges E, et al. Updates to the RMAP short-read mapping software. *Bioinformatics*. 2009;25:2841–2.
164. Otto C, Stadler PF, Hoffmann S. Fast and sensitive mapping of bisulfite-treated sequencing data. *Bioinformatics*. 2012;28:1698–704.
165. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics*. 2011;27:1571–2.
166. Harris EY, Ponts N, Levchuk A, Roch KL, Lonardi S. BRAT: bisulfite-treated reads analysis tool. *Bioinformatics*. 2010;26:572–3.
167. Chen PY, Cokus SJ, Pellegrini M. BS seeker: precise mapping for bisulfite sequencing. *BMC Bioinformatics*. 2010;11:203.
168. Pedersen B, Hsieh TF, Ibarra C, Fischer RL. MethylCoder: software pipeline for bisulfite-treated sequences. *Bioinformatics*. 2011;27:2435–6.
169. Xi Y, Li W. BSMAP: whole genome bisulfite sequence mapping program. *BMC Bioinformatics*. 2009;10:232.
170. Coarfa C. Pash 3.0: a versatile software package for read mapping and integrative analysis of genomic and epigenomic variation using massively parallel DNA sequencing. *BMC Bioinformatics*. 2011;11:572.
171. Coarfa C, Milosavljevic A. Pash 2.0: scaleable sequence anchoring for next-generation sequencing technologies. *Pac Symp Biocomput*. 2008:102–13.
172. Kalafus KJ, Jackson AR, Milosavljevic A. Pash: efficient genome-scale sequence anchoring by positional hashing. *Genome Res*. 2004;14(4):672–8.
173. Xi Y, Bock C, Müller F, Sun D, Meissner A, Li W. RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing. *Bioinformatics*. 2012;28:430–2.
174. Wang D, Yan L, Hu Q, et al. IMA: an R package for high-throughput analysis of illumina's 450 K infinium methylation data. *Bioinformatics*. 2012;28(5):729–30.
175. Hebestreit K, Dugas M, Klein HU. Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. *Bioinformatics*. 2013;29(13):1647–53.

176. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol*. 2012;13(10):R83.
177. Jaffe AE, Murakami P, Lee H, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol*. 2012;41(1):200–9.
178. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015;8:6.
179. Mayo TR, Schweikert G, Sanguinetti G. M3D: a kernel-based test for spatially correlated changes in methylation profiles. *Bioinformatics*. 2015;31(6):809–16.
180. Park Y, Figueroa ME, Rozek LS, Sartor MA. MethylSig: a whole genome DNA methylation analysis pipeline. *Bioinformatics*. 2014;30(17):2414–22.
181. Sun D, Xi Y, Rodriguez B, et al. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol*. 2014;15(2):R38.
182. Kuan PF, Chiang DY. Integrating prior knowledge in multiple testing under dependence with applications to detecting differential DNA methylation. *Biometrics*. 2012;68(3):774–83.
183. Song Q, Decato B, Hong EE, et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One*. 2013;8(12):e81148.
184. Zhang Y, Liu H, Lv J, et al. QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res*. 2011;39(9):e58.
185. Juhling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res*. 2016;26(2):256–62.
186. Booth MJ, Branco MR, Ficz G, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*. 2012;336(6083):934–7.
187. Yu M, Hon GC, Szulwach KE, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*. 2012;149(6):1368–80.
188. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res*. 2013;23(12):2126–35.
189. Smallwood SA, Lee HJ, Angermueller C, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods*. 2014;11(8):817–20.
190. Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 2010;7(6):461–5.
191. Laszlo AH, Derrington IM, Brinkerhoff H, et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A*. 2013;110(47):18904–9.
192. Schreiber J, Wescoe ZL, Abu-Shumays R, et al. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc Natl Acad Sci U S A*. 2013;110(47):18910–5.
193. Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD. Genomic landscape of human allele-specific DNA methylation. *Proc Natl Acad Sci U S A*. 2012;109(19):7332–7.
194. Peng Q, Ecker JR. Detection of allele-specific methylation through a generalized heterogeneous epigenome model. *Bioinformatics*. 2012;28(12):i163–71.
195. Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Hum Mol Genet*. 2007;16(5):547–54.
196. Bjornsson HT, Sigurdsson MI, Fallin MD, et al. Intra-individual change over time in DNA methylation with familial clustering. *JAMA*. 2008;299(24):2877–83.
197. Bock KW. Functions and transcriptional regulation of adult human hepatic UDP-glucuronosyl-transferases (UGTs): mechanisms responsible for interindividual variation of UGT levels. *Biochem Pharmacol*. 2010;80(6):771–7.
198. Gertz J, Varley KE, Reddy TE, et al. Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet*. 2011;7(8):e1002228.
199. Ji H, Ehrlich LI, Seita J, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*. 2010;467(7313):338–42.
200. Bock C, Beerman I, Lien WH, et al. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol Cell*. 2012;47(4):633–47.
201. Hansen KD, Timp W, Bravo HC, et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*. 2011;43(8):768–75.
202. Zheng X, Zhao Q, Wu HJ, et al. MethylPurify: tumor purity deconvolution and differential methylation detection from single tumor DNA methylomes. *Genome Biol*. 2014;15(8):419.
203. Zhang N, Wu HJ, Zhang W, Wang J, Wu H, Zheng X. Predicting tumor purity from methylation microarray data. *Bioinformatics*. 2015;31(21):3401–5.
204. Stirzaker C, Taberlay PC, Statham AL, Clark SJ. Mining cancer methylomes: prospects and challenges. *Trends Genet*. 2014;30(2):75–84.
205. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011;12(8):529–41.
206. Laird PW. The power and the promise of DNA methylation markers. *Nat Rev Cancer*. 2003;3(4):253–66.
207. Glockner SC, Dhir M, Yi JM, et al. Methylation of TFPI2 in stool DNA: a potential novel biomarker for the detection of colorectal cancer. *Cancer Res*. 2009;69(11):4691–9.
208. Lofton-Day C, Model F, Devos T, et al. DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin Chem*. 2008;54(2):414–23.
209. Zamani M, Hosseini SV, Mokarram P. Epigenetic biomarkers in colorectal cancer: premises and prospects. *Biomarkers*. 2016;28:1–38.
210. Cairns P, Esteller M, Herman JG, et al. Molecular detection of prostate cancer in urine by GSTP1 hypermethylation. *Clin Cancer Res*. 2001;7(9):2727–30.
211. Rosenbaum E, Hoque MO, Cohen Y, et al. Promoter hypermethylation as an independent prognostic factor for relapse in patients with prostate cancer following radical prostatectomy. *Clin Cancer Res*. 2005;11(23):8321–5.
212. Zhao F, Olkhov-Mitsel E, van der Kwast T, et al. Urinary DNA methylation biomarkers for noninvasive prediction of aggressive disease in patients with prostate cancer on active surveillance. *J Urol*. 2016:16:31067–9.
213. Brock MV, Hooker CM, Ota-Machida E, et al. DNA methylation markers and early recurrence in stage I lung cancer. *N Engl J Med*. 2008;358(11):1118–28.
214. Su Y, Fang H, Jiang F. Integrating DNA methylation and microRNA biomarkers in sputum for lung cancer detection. *Clin Epigenetics*. 2016;8:109.
215. Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell*. 2007;128(4):707–19.
216. Santos-Rosa H, Kirmizis A, Nelson C, et al. Histone H3 tail clipping regulates gene expression. *Nat Struct Mol Biol*. 2009;16(1):17–22.
217. Wang Z, Zang C, Rosenfeld JA, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*. 2008;40(7):897–903.
218. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*. 2010;28(8):817–25.
219. Valouev A, Johnson DS, Sundquist A, et al. Genome-wide analysis of transcription factor binding sites based on ChIP-seq data. *Nat Methods*. 2008;5(9):829–34.
220. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol*. 2008;26(12):1351–9.
221. Adli M, Bernstein BE. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nat Protoc*. 2011;6(10):1656–68.
222. Shankaranarayanan P, Mendoza-Parra MA, Walia M, et al. Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nat Methods*. 2011;8(7):565–7.
223. Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*. 2011;147(6):1408–19.
224. Furlan-Magaril M, Rincon-Arano H, Recillas-Targa F. Sequential chromatin immunoprecipitation protocol: ChIP-reChIP. *Methods Mol Biol*. 2009;543:253–66.
225. Brinkman AB, Gu H, Bartels SJ, et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res*. 2012;22(6):1128–38.
226. Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, Clark SJ. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res*. 2012;22(6):1120–7.
227. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform*. 2010;11(5):473–83.
228. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
229. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics*. 2008;24(5):713–4.
230. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 2008;18(11):1851–8.
231. Lin H, Zhang Z, Zhang MQ, Ma B, Li M. ZOOM! Zillions of oligos mapped. *Bioinformatics*. 2008;24(21):2431–7.
232. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
233. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
234. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009;25(15):1966–7.
235. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. *Nat Methods*. 2009;6(suppl 11):S22–32.
236. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-seq data. *Bioinformatics*. 2009;25(15):1952–8.
237. Wang J, Lunyak VV, Jordan IK. BroadPeak: a novel algorithm for identifying broad peaks in diffuse ChIP-seq datasets. *Bioinformatics*. 2013;29(4):492–3.
238. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol*. 2008;9(9):R137.
239. Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-seq data. *Nucleic Acids Res*. 2008;36(16):5221–31.

240. Wilbanks EG, Facciotti MT. Evaluation of algorithm performance in ChIP-seq peak detection. *PLoS One*. 2010;5(7):e11471.
241. Steinhauser S, Kurzawa N, Eils R, Herrmann C. A comprehensive comparison of tools for differential ChIP-seq analysis. *Brief Bioinform*. 2016.
242. Tordonato C, Di Fiore PP, Nicassio F. The role of non-coding RNAs in the regulation of stem cells and progenitors in the normal mammary gland and in breast tumors. *Front Genet*. 2015;6:72.
243. Necsulea A, Kaessmann H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet*. 2014;15(11):734–48.
244. Peschansky VJ, Wahlestedt C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics*. 2014;9(1):3–12.
245. Zhou H, Hu H, Lai M. Non-coding RNAs and their epigenetic regulatory mechanisms. *Biol Cell*. 2010;102(12):645–55.
246. Amaral PP, Mattick JS. Noncoding RNA in development. *Mamm Genome*. 2008;19(7–8):454–92.
247. Lu ZJ, Yip KY, Wang G, et al. Prediction and characterization of noncoding RNAs in *C. elegans* by integrating conservation, secondary structure, and high-throughput sequencing and array data. *Genome Res*. 2011;21(2):276–85.
248. Friedlander MR, Chen W, Adamidi C, et al. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol*. 2008;26(4):407–15.
249. An J, Lai J, Lehman ML, Nelson CC. miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res*. 2013;41(2):727–37.
250. Giurato G, De Filippo MR, Rinaldi A, et al. iMir: an integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-seq. *BMC Bioinformatics*. 2013;14:362.
251. Hackenberg M, Rodriguez-Ezpeleta N, Aransay AM. miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Res*. 2011;39(Web Server issue):W132–8.
252. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res*. 2012;40(1):37–52.
253. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.
254. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015;4.
255. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res*. 2008;36(Database issue):D149–53.
256. Sun Z, Evans J, Bhagwate A, et al. CAP-miRSeq: a comprehensive analysis pipeline for microRNA sequencing data. *BMC Genomics*. 2014;15:423.
257. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
258. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303.
259. Muller S, Rycak L, Winter P, Kahl G, Koch I, Rotter B. omiRas: a web server for differential expression analysis of miRNAs derived from small RNA-seq data. *Bioinformatics*. 2013;29(20):2651–2.
260. Stocks MB, Moxon S, Mapleson D, et al. The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*. 2012;28(15):2059–61.
261. Sun K, Chen X, Jiang P, Song X, Wang H, Sun H. iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data. *BMC Genomics*. 2013;14(suppl 2):S7.
262. Sun K, Zhao Y, Wang H, Sun H. Sebnif: an integrated bioinformatics pipeline for the identification of novel large intergenic noncoding RNAs (lincRNAs) – application in human skeletal muscle cells. *PLoS One*. 2014;9(1):e84500.
263. Li G, Pan W, Yang X, Miao J. Gene co-expression network and function modules in three types of glioma. *Mol Med Rep*. 2015;11(4):3055–63.
264. Jiang Q, Ma R, Wang J, et al. LncRNA2Function: a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data. *BMC Genomics*. 2015;16(suppl 3):S2.
265. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res*. 2012;22(9):1760–1774.
266. Uren PJ, Bahrami-Samani E, Burns SC, et al. Site identification in high-throughput RNA-protein interaction data. *Bioinformatics*. 2012;28(23):3013–20.
267. Krawitz P, Rodelsperger C, Jager M, Jostins L, Bauer S, Robinson PN. Microindel detection in short-read sequence data. *Bioinformatics*. 2010;26(6):722–9.
268. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics*. 2009;25(9):1105–11.
269. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
270. Chen B, Yun J, Kim MS, Mendell JT, Xie Y. PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. *Genome Biol*. 2014;15(1):R18.
271. Corcoran DL, Georgiev S, Mukherjee N, et al. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol*. 2011;12(8):R79.
272. Sievers C, Schlumpf T, Sawarkar R, Comoglio F, Paro R. Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. *Nucleic Acids Res*. 2012;40(20):e160.
273. Wang T, Xie Y, Xiao G. dCLIP: a computational approach for comparative CLIP-seq analyses. *Genome Biol*. 2014;15(1):R11.
274. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89.
275. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37(Web Server issue):W202–8.
276. Georgiev S, Boyle AP, Jayasurya K, Ding X, Mukherjee S, Ohler U. Evidence-ranked motif identification. *Genome Biol*. 2010;11(2):R19.
277. Frith MC, Saunders NF, Kobe B, Bailey TL. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol*. 2008;4(4):e1000071.
278. Foat BC, Morozov AV, Bussemaker HJ. Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics*. 2006;22(14):e141–9.
279. Kazan H, Ray D, Chan ET, Hughes TR, Morris Q. RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. *PLoS Comput Biol*. 2010;6:e1000832.
280. Maticzka D, Lange SJ, Costa F, Backofen R. GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol*. 2014;15(1):R17.
281. Fukunaga T, Ozaki H, Terai G, Asai K, Iwasaki W, Kiryu H. CapR: revealing structural specificities of RNA-binding protein target recognition using CLIP-seq data. *Genome Biol*. 2014;15(1):R16.
282. Rosenbloom KR, Armstrong J, Barber GP, et al. The UCSC genome browser database: 2015 update. *Nucleic Acids Res*. 2015;43(Database issue):D670–81.
283. Yates A, Akanni W, Amode MR, et al. Ensembl 2016. *Nucleic Acids Res*. 2016;44(D1):D710–6.
284. Zhou X, Maricque B, Xie M, et al. The human epigenome browser at Washington University. *Nat Methods*. 2011;8(12):989–90.
285. Fejes AP, Jones MJ, Kobor MS. DaVIE: database for the visualization and integration of epigenetic data. *Front Genet*. 2014;5:325.
286. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24–6.
287. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets – update. *Nucleic Acids Res*. 2013;41(Database issue):D991–5.
288. Grunau C, Renault E, Rosenthal A, Roizes G. MethDB – a public database for DNA methylation data. *Nucleic Acids Res*. 2001;29:270–4.
289. He X, Chang S, Zhang J, et al. MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Res*. 2008;36(Database issue):D836–41.
290. Ongenaert M, Van Neste L, De Meyer T, Menschaert G, Bekaert S, Van Criekinge W. PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucleic Acids Res*. 2008;36(Database issue):D842–6.
291. Zhang Y, Lv J, Liu H, et al. HHMD: the human histone modification database. *Nucleic Acids Res*. 2010;38(Database issue):D149–54.
292. Khare SP, Habib F, Sharma R, Gadewal N, Gupta S, Galande S. HIstome – a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res*. 2012;40(Database issue):D337–42.
293. Wang Q, Huang J, Sun H, et al. CR cistrome: a ChIP-seq database for chromatin regulators and histone modification linkages in human and mouse. *Nucleic Acids Res*. 2014;42(Database issue):D450–8.
294. O'Connor TR, Wyrick JJ. ChromatinDB: a database of genome-wide histone modification patterns for *Saccharomyces cerevisiae*. *Bioinformatics*. 2007;23(14):1828–30.
295. Rustici G, Kolesnikov N, Brandizi M, et al. ArrayExpress update – trends in database growth and links to data analysis tools. *Nucleic Acids Res*. 2013;41(Database issue):D987–90.
296. Yang JH, Li JH, Shao P, Zhou H, Chen YQ, Qu LH. starBase: a database for exploring microRNA-mRNA interaction maps from argonaute CLIP-seq and degradome-seq data. *Nucleic Acids Res*. 2011;39(Database issue):D202–9.
297. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92–7.
298. Khorshid M, Rodak C, Zavolan M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res*. 2011;39(Database issue):D245–52.
299. Blin K, Dieterich C, Wurmus R, Rajewsky N, Landthaler M, Akalin A. DoRiNA 2.0 – upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res*. 2015;43(Database issue):D160–7.
300. Anders G, Mackowiak SD, Jens M, et al. doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res*. 2012;40(Database issue):D180–6.
301. Yang YC, Di C, Hu B, et al. CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics*. 2015;16:51.
302. Adams D, Altucci L, Antonarakis SE, et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol*. 2012;30(3):224–6.
303. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol*. 2010;28(10):1045–8.

304. ENCODE Project Consortium, Dunham I, Kundaje A, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
305. Harrow J, Denoeud F, Frankish A, et al. GENCODE: producing a reference annotation for ENCODE. *Genome Biol*. 2006;7(suppl 1):S41–9.
306. Ravasi T, Suzuki H, Pang KC, et al. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res*. 2006;16(1):11–9.
307. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell*. 2013;153(1):17–37.
308. International Cancer Genome Consortium, Hudson TJ, Anderson W, et al. International network of cancer genome projects. *Nature*. 2010;464(7291):993–8.
309. Liu Y, Han D, Han Y, et al. Ab initio identification of transcription start sites in the Rhesus macaque genome by histone modification and RNA-seq. *Nucleic Acids Res*. 2011;39(4):1408–18.
310. Roadmap Epigenomics C, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317–30.
311. Balbin OA, Prensner JR, Sahu A, et al. Reconstructing targetable pathways in lung cancer by integrating diverse omics data. *Nat Commun*. 2013;4:2617.
312. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*. 2012;9(3):215–6.
313. Choi H, Fermin D, Nesvizhskii AI, Ghosh D, Qin ZS. Sparsely correlated hidden Markov models with application to genome-wide location studies. *Bioinformatics*. 2013;29(5):533–41.
314. Yu H, Zhu S, Zhou B, Xue H, Han JD. Inferring causal relationships among different histone modifications and gene expression. *Genome Res*. 2008;18(8):1314–24.
315. Liu Y, Qiao N, Zhu S, et al. A novel Bayesian network inference algorithm for integrative analysis of heterogeneous deep sequencing data. *Cell Res*. 2013;23(3):440–3.
316. Lasserre J, Chung HR, Vingron M. Finding associations among histone modifications using sparse partial correlation networks. *PLoS Comput Biol*. 2013;9(9):e1003168.
317. Shen L, Toyota M, Kondo Y, et al. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A*. 2007;104(47):18654–9.
318. Heintzman ND, Stuart RK, Hon G, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007;39(3):311–8.
319. Mo Q, Wang S, Seshan VE, et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc Natl Acad Sci U S A*. 2013;110(11):4245–50.
320. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513(7517):202–9.
321. Konig J, Zarnack K, Rot G, et al. iCLIP – transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution. *J Vis Exp*. 2011;50:2638.
322. Hafner M, Landthaler M, Burger L, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010;141(1):129–41.
323. Clement NL, Snell Q, Clement MJ, et al. The GNUMAP algorithm: unbiased probabilistic mapping of oligonucleotides from next-generation sequencing. *Bioinformatics*. 2010;26(1):38–45.
324. Barash Y, Calarco JA, Gao W, et al. Deciphering the splicing code. *Nature*. 2010;465(7294):53–9.
325. Agger K, Cloos PA, Rudkjaer L, et al. The H3 K27 me3 demethylase JMJD3 contributes to the activation of the INK4 A-ARF locus in response to oncogene- and stress-induced senescence. *Genes Dev*. 2009;23(10):1171–6.
326. Rohde C, Zhang Y, Reinhardt R, Jeltsch A. BISMA – fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. *BMC Bioinformatics*. 2010;11:230.
327. Kreck B, Marnellos G, Richter J, Krueger F, Siebert R, Franke A. B-SOLANA: an approach for the analysis of two-base encoding bisulfite sequencing data. *Bioinformatics*. 2012;28(3):428–9.
328. BasuRay S, Mukherjee S, Romero EG, Seaman MN, Wandinger-Ness A. Rab7 mutants associated with Charcot-Marie-tooth disease cause delayed growth factor receptor transport and altered endosomal and nuclear signaling. *J Biol Chem*. 2013;288(2):1135–49.
329. Chen KC, Liao YC, Hsieh IC, Wang YS, Hu CY, Juo SH. OxLDL causes both epigenetic modification and signaling regulation on the microRNA-29b gene: novel mechanisms for cardiovascular diseases. *J Mol Cell Cardiol*. 2012;52(3):587–95.
330. Gaisina IN, Tueckmantel W, Ugolkov A, et al. Identification of HDAC6-selective inhibitors of low cancer cell cytotoxicity. *ChemMedChem*. 2016;11(1):81–92.
331. Jang W, Park HH, Lee KY, Lee YJ, Kim HT, Koh SH. 1,25-dyhydroxyvitamin D3 attenuates L-DOPA-induced neurotoxicity in neural stem cells. *Mol Neurobiol*. 2015;51(2):558–70.
332. Farioli-Vecchioli S, Ceccarelli M, Saraulli D, et al. Tis21 is required for adult neurogenesis in the subventricular zone and for olfactory behavior regulating cyclins, BMP4, Hes1/5 and Ids. *Front Cell Neurosci*. 2014;8:98.
333. Gui Y, Guo G, Huang Y, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet*. 2011;43(9):875–8.
334. Liu WS, Zhao LJ, Pang QS, Yuan ZY, Li B, Wang P. Prognostic value of epidermal growth factor receptor mutations in resected lung adenocarcinomas. *Med Oncol*. 2014;31(1):771.
335. Klughammer J, Datlinger P, Printz D, et al. Differential DNA methylation analysis without a reference genome. *Cell Rep*. 2015;13(11):2621–33.
336. Schneeberger K, Hagmann J, Ossowski S, et al. Simultaneous alignment of short reads against multiple genomes. *Genome Biol*. 2009;10(9):R98.
337. Wingett S, Ewels P, Furlan-Magaril M, et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res*. 2015;4:1310.
338. Rozowsky J, Euskirchen G, Auerbach RK, et al. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol*. 2009;27(1):66–75.
339. Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD. ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. *Genome Biol*. 2011;12(7):R67.
340. Hardcastle TJ, Kelly KA. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*. 2010;11:422.
341. Xu H, Wei CL, Lin F, Sung WK. An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. *Bioinformatics*. 2008;24(20):2344–9.
342. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-seq data. *Stat Methods Med Res*. 2013;22(5):519–36.
343. Fasold M, Langenberger D, Binder H, Stadler PF, Hoffmann S. DARIO: a ncRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res*. 2011;39(Web Server issue):W112–7.
344. Chen CJ, Servant N, Toedling J, et al. ncPRO-seq: a tool for annotation and profiling of ncRNAs in sRNA-seq data. *Bioinformatics*. 2012;28(23):3147–9.
345. Leung YY, Ryvkin P, Ungar LH, Gregory BD, Wang LS. CoRAL: predicting non-coding RNAs from small RNA-sequencing data. *Nucleic Acids Res*. 2013;41(14):e137.
346. Yuan C, Sun Y. RNA-CODE: a noncoding RNA classification tool for short reads in NGS data lacking reference genomes. *PLoS One*. 2013;8(10):e77596.
347. Rueda A, Barturen G, Lebron R, et al. sRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic Acids Res*. 2015;43(W1):W467–73.
348. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*. 1994;2:28–36.
349. Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB. RNA bind-n-seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell*. 2014;54(5):887–900.
350. Rohde C, Zhang Y, Jurkowski TP, Stamerjohanns H, Reinhardt R, Jeltsch A. Bisulfite sequencing data presentation and compilation (BDPC) web server – a useful tool for DNA methylation analysis. *Nucleic Acids Res*. 2008;36(5):e34.
351. 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.
352. Bock C, Halachev K, Buch J, Lengauer T. EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data. *Genome Biol*. 2009;10(2):R14.
353. Guan Z, Xu B, DeSilvio ML, et al. Randomized trial of lapatinib versus placebo added to paclitaxel in the treatment of human epidermal growth factor receptor 2-overexpressing metastatic breast cancer. *J Clin Oncol*. 2013;31(16):1947–53.
354. Proceedings of the First Russian-European Workshop on DNA Repair and Epigenetic Regulation of Genome Stability. June 24–6, 2008. St. Petersburg, Russia. Dedicated to the Memory of Nikolay Tomilin. *Mutat Res*. 2010;685(1–2):1–102.