

Vladimir Espinosa Angarica and Antonio del Sol

Abstract

Epigenetics play a central role in the regulation of many important cellular processes, and dysregulations at the epigenetic level could be the source of serious pathologies, such as neurological disorders affecting brain development, neurodegeneration, and intellectual disability. Despite significant technological advances for epigenetic profiling, there is still a need for a systematic understanding of how epigenetics shapes cellular circuitry, and disease pathogenesis. The development of accurate computational approaches for analyzing complex epigenetic profiles is essential for disentangling the mechanisms underlying cellular development, and the intricate interaction networks determining and sensing chromatin modifications and DNA methylation to control gene expression. In this chapter, we review the recent advances in the field of “computational epigenetics,” including computational methods for processing different types of epigenetic data, prediction of chromatin states, and study of protein dynamics. We also discuss how “computational epigenetics” has complemented the fast growth in the generation of epigenetic data for uncovering the main differences and similarities at the epigenetic level between individuals and the mechanisms underlying disease onset and progression.

Keywords

Computational epigenetics • Histone code • Epigenetics regulation • Transcriptional regulation • Next generation sequencing analysis • Single cell neuroepigenetics

V.E. Angarica, Ph.D. (✉) • A. del Sol, Ph.D.
Computational Biology Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 Avenue du Swing, 4366 Belvaux, Luxembourg
e-mail: antonio.delsol@uni.lu

25.1 Chromatin Structure, Combinatorial Complexity of Histone Modifications, and Mechanisms of Epigenetic Regulation

Epigenetic phenomena constitute a very important regulatory checkpoint in many key cellular processes such as DNA maintenance and repair [1, 2], epigenetic inheritance [3, 4], and gene expression [5, 6]. While the genome underlying structure – i.e., DNA sequence – is highly stable, epigenetic signatures are dynamic [7–9], with different epigenetic phenomena having different degrees of stability and variability, causing most of the phenotypic differences across cells in multicellular organisms. Fluctuations in DNA condensation, and the establishment of heterochromatic or euchromatic regions, are determined by covalent modifications of chromatin, including DNA methylation of CpG islands [10–12], and a wide range of histone modifications [9, 13, 14], which form complex combinatorial networks of histone marks, that constitute the “histone code” [15]. Moreover, DNA methylation and histone modification pathways are significantly interconnected [16–18], and the cross talk between DNA and histone epigenetic modifications significantly increases the combinatorial complexity of the mechanisms of epigenetic regulation. Although not yet fully understood, there are two characterized mechanisms by which epigenetic modifications exert their function [9]: the first is the disruption of contacts between nucleosomes in order to “unravel” chromatin, and the second is the recruitment of nonhistone proteins [9]. A wide family of epigenetic signaling proteins – i.e., readers, writers, and erasers – [19–22] recognize the complex code of epigenetic modifications, controlling the condensation levels of genomic regions, and the susceptibility of these regions to be transcribed [5, 6], to be subject of DNA repair [1, 2] or be involved in other cellular processes. The central role of epigenetics in the regulation of a broad range of key cellular processes explains their implication in multiple common and serious human pathologies [23–25], such as developmental diseases [26–28], cancer [29–32], and neurological disorders [33–37]. Despite technological advances for the study of mechanisms of epigenetic regulation, we still lack a systematic understanding of how the epigenomic landscape contributes to cellular circuitry, lineage specification, and the onset and progression of human disease [38]. Due to the significant complexity of the mechanisms of epigenetic regulation, computational and bioinformatics approaches have been essential for disentangling these mechanisms at the genome-wide level and for answering important questions such as how the epigenetic level senses environmental cues during lineage specification and development and which are the interactions among different chromatin modifications to control transcription.

In this chapter, we review the state of the art of computational approaches and bioinformatics tools for genome-wide epigenetic research. We cover the field of “computational epigenetics” and discuss recent advances in computational methods for processing and quality control of different types of epigenetic data, the prediction of chromatin states and the study of the dynamics of chromatin, and the analysis of 3D structure of chromatin. We also address the status of different collaborative projects and databases comprising a wealth of genome-wide epigenetic data. We discuss how the fast growth in the generation of epigenetic data, boosted by the development of

high-throughput sequencing (HTS) experimental technologies and inter-institutional public/private collaborative projects, has been complemented and prompted by the development of computational methods for analyzing and rationalizing huge quantities of epigenetic data. The steady decrease in the cost of technologies for generating epigenetic data has also opened the possibility of performing epigenetic surveys in human populations. In this regard, we also examine the recent development in the computational approaches used to perform these studies for uncovering the main differences and similarities at the epigenetic level between individuals and their implication in cellular differentiation, gene regulation, and disease.

25.2 Whole Genome Annotation of Histone Modifications: Computational Tools for Data Quality Control and Mapping of Epigenetic Data

The characteristics and specificities of the wide range of computational methods commonly used for the analysis of epigenetic data depend significantly on the particularities of the experimental techniques used to perform epigenomic profiling. The techniques available for profiling histone modifications (and the other epigenetic phenomena described in the next sections of this chapter) are described in detail in a previous chapter of this book, but it would be important to summarize their commonalities and differences to discuss the different computational approaches used to analyze the epigenetic data generated in each case. The most commonly used experimental approaches to profile histone posttranslational modifications are ChIP-on-chip [39–41], ChIP-seq [42–44], and mass spectrometry [45–48]. In ChIP-on-chip, histone modification-specific antibodies, bound to chromatin regions bearing the corresponding modification, are cross-linked to DNA by treatment with formaldehyde. Next, chromatin is collected and fragmented by sonication or using nucleases, and the fragments bearing the histone modification are enriched by using an antibody matrix specific to the histone modification-specific antibody – i.e., immunoprecipitation. The DNA in the enriched fragments is released reverting cross-linking by increasing temperature, and purified DNA fragments are amplified and labeled with fluorescent dyes for further quantitation. Finally, purified DNA is hybridized to a tiling microarray, which allows the identification of regions over-represented in the immunoprecipitated DNA relative to control DNA – i.e., regarded as epigenetically modified. ChIP-seq shares the initial steps of the ChIP-on-chip technique, but unlike the former, it relies on HTS DNA sequencing rather than on microarrays for identifying the sequences enriched in histone marks. Unlike immunoprecipitation techniques, proteomic profiling using mass spectrometry (MS) allows the detailed characterization of histone tail posttranslational modifications. This technique relies on the chromatographic separation of histones from cell lysates, followed by enzymatic digestion of individual histones for the accurate assignment and quantification of the amino acids bearing different kinds of post-translational modifications [9, 13, 14], following top-down, bottom-up, or middle-down approaches [47, 49].

Immunoprecipitation techniques are by far the most commonly used, thanks to their high-throughput capabilities and the developments in the production of highly specific histone modification-specific antibodies. The main bioinformatics problem for the analysis of ChIP-on-chip data is establishing a ranking of genomic regions overrepresented on the arrays from raw probe intensities. In this regard, many different approaches have been specifically developed for performing peak calling from ChIP-on-chip experiments. In general, these methods have a set of common steps, encompassing the normalization of the intensities of hybridized fragments, assessment of the statistical significance of the intensities of each peak with respect to the whole array, and finally merging overlapping overrepresented regions [39–41, 50]. The list of peak-calling packages for processing ChIP-on-chip data is fairly ample and diverse, including Telescope [51], an automated data processing toolkit for analyzing high-density tiling microarray data that integrates data normalization, combination of replicate experiments, tile scoring, and feature identification in an easy-to-use online suite. Tilemap [52] is a stand-alone package that provides a flexible way to study tiling array hybridizations under multiple experimental conditions in Affymetrix ChIP-on-chips. Ringo [53] is an R package devised for NimbleGen microarrays, which facilitates the construction of automated programmed workflows and enables the scalability and reproducibility of the analyses in comparison to other ChIP-on-chip peak callers. The abovementioned list of bioinformatics tools for processing ChIP-on-chip microarray data is by no means exhaustive, and there is a wide spectrum of other approaches, including ACME [54], HGMM [55], ChIPOTle [56], HMMTiling [57], and MAT [58], among others. Notwithstanding the diversity of tools for processing ChIP-on-chip data, the bioinformatics analysis of tiling microarrays shares the same drawbacks of the algorithms for analyzing DNA arrays, as they fail to accurately estimate histone modifications spanning extended genomic regions and underestimate weak binding events [50].

The key bioinformatics challenge in the analysis of ChIP-seq data is the fast and accurate mapping of thousands to millions of short reads, corresponding to the regions bearing a specific histone modification, to the reference genome. Many sequence aligners for solving the problems of mapping short sequence reads have been developed, such as Bowtie [59], BWA [60], SOAP [61], and BLAT [62], among a wide list of others (for a detailed review on short-read alignment methods see [63]). Other methods with alignment strategies optimized for reads obtained with specific sequencing platforms have been developed, including commercial suites such as ELAND that form part of the SOLEXA pipeline (<http://www.solexa.com/>), and the Broad Institute sequencing platform [64] (<http://genomics.broadinstitute.org/>). While mapping short reads to a reference genome, special care should be taken to the quality control of sequencing data. For instance, random fragmentation of ChIP-seq samples treated with sonication renders an array of overlapping reads corresponding to the same genomic regions, and these duplicated reads should be removed, using for example SAMtools [65]. This requirement for quality control is not necessary, however, while analyzing ChIP-seq data generated from samples treated with nucleases, because the likelihood of the generation of overlapping reads is rather low. The assessment of “uniquely mapped” and “unique reads” is also a very important step in

the quality control of ChIP-seq data. The former correspond to reads that aligned to specific regions, excluding repetitive genomic *loci* and non-repetitive regions with highly similar sequences, while the latter correspond to de-duplication PCR reads. In this regard, depending on the specificities of the ChIP-seq dataset, removal of duplicated reads to reduce amplification artifacts could result in an underestimation of real binding events. On the other hand, not removing duplicate reads could cause the inclusion of a significant amount of false positives, which could have strong implications in the downstream analysis of ChIP-seq data. Therefore, alignment of short sequence reads to the reference genome, and quality control of sequencing data, still remains a bioinformatics challenge. The analysis of the signal-to-noise ratio of sequencing signals also constitutes an important step on ChIP-seq quality control. The estimation of the “fraction of reads in peaks” (FRiP) – i.e., number of reads per region – and cross-correlation profiles (CCPs), i.e., read clustering prior to peak calling [66], are very useful for assessing the signal-to-noise ratio. Based on these metrics, different approaches for estimating the signal-to-noise ratio of ChIP-seq sequencing data have been developed [67].

The procedures for performing peak calling from ChIP-seq samples are different from those commonly used for ChIP-on-chip experiments. There exists a myriad of different peak callers based on different statistical criteria, which cannot be covered here in detail (for a detailed review, please see [68]). The general procedure followed by all of these algorithms includes the identification of enriched sequence read density for different chromosome *loci*, relative to a background sequence read distribution. The first step common to all ChIP-seq peak callers is the generation of a signal profile by integrating reads mapped to specific genomic regions. Different tools rely on sliding-window approaches for smoothing the discrete distribution of read counts into a continuous signal profile distribution. Tools such as CisGenome [69] follow this rationale, estimating the number of reads above a predefined peak cutoff, and others like SISR [70], Peakzilla [71], and SPP [72] also take into account the correspondence of read counts in positive and negative strands to improve peak resolution. Other tools use more sophisticated approaches for integrating the signals in sequence windows. For example, MACS [73] uses the local Poisson model to identify local biases in genomic positions, F-Seq [74] and QuEST [75] rely on kernel density estimations, and PICS [76] uses a Bayesian hierarchical *t*-mixture model for smoothing count reads in the genomic signal profile. The HOMER program suite [77] has been also widely used for peak calling and is specially useful for analyzing broad peak corresponding to histone modifications – e.g., H3K9me3 – spanning large chromosome regions. Other tools such as JAMM [78] and PePr [79] integrate information from biological replicates to determine enrichment site widths in neighboring narrow peaks, whereas GLITR [80] and PeakSeq [81] use tag extension –i.e., extension of ChIP-seq tags along their strand direction – to identify genomic regions enriched in sequence reads. The selection of the background distribution used in the comparison with the sample analyzed is also an essential step in peak calling. Although there is no consensus on which is the best background distribution, different datasets have been used as control sample, such as ChIP-seq data for histone H3, or from experiments using a control antibody for

nonbinding proteins, such as immunoglobulins [66, 82]. The following steps during peak calling include the selection of the statistical criteria for identifying enriched peaks, which usually correspond to a specific cutoff for the enrichment of peaks relative to the background, or estimating metrics with more statistical support, such as the false discovery rate (FDR). Once enriched peaks are identified for a selected number of genes, or genome wide, most peak calling algorithms allow ranking and selection of the more significant peaks by estimating their corresponding p -values and q -values. Despite the great variety of peak calling toolkits for analyzing ChIP-seq data, the comparison of the performance of different approaches shows that different programs produce very different peaks in terms of peak size, number, and position relative to genes [83, 84] when presented to the same input dataset. Thus, as different tools usually generate significantly different epigenomic profiles, peak calling of ChIP-seq data remains a difficult task, and the selection of the best performing methods usually depends on the species, sample conditions, and target proteins [43].

The bioinformatics analysis of histone posttranslational modification profiles obtained with MS is significantly dependent on the specific MS approach used – e.g., top-down, bottom-up, or middle-down approaches [47, 49]. The preprocessing of MS data for removing false fragment ion assignments can be performed with different programs, such as Thrash [85], MS-Deconv [86], or YADA [87]. These approaches can also be used to deconvolute ion signals with multiple charges into mono-charged ion mass values from bottom-up MS profiles, but are unable to produce good results for other approaches generating longer peptides [88]. Unlike immunoprecipitation techniques, in which PTM-specific antibodies are used to profile one histone modification per experiment, the analysis of cell lysates with MS has the added difficulty of having to deal with the genome-wide profile of all the histone modifications. Due to the huge combinatorial complexity of this problem, current approaches concentrate on the most common histone PTM [47], which might overlook unknown, but functionally relevant modifications. Top-down and middle-down proteomics strategies require specialized search algorithms and annotation tools, due to the great complexity of the MS spectra generated for intact or large polypeptides [89]. Methods such as ProSight PTM [90], MX-Align+ [91], ROCCIT [92], and MLIP [93] are tools specifically suited for performing database sequence searches from neutral mass lists of precursor and fragment ions obtained with top-down approaches. Different implementations of the THRASH [85] algorithm have been adapted for top-down histone modification profiling [94, 95], as well as MS-Deconv tool [86], developed specifically to analyze MS spectra from complete proteins. These methods offer a number of different functionalities for guiding the search for specific modifications that allows a significant reduction of the search space, which can increase the significance of assigned peaks. Other tools allow tackling the complex problem of identifying different histone PTM fragments with fairly similar ion masses [93, 96]. The software VEMS is included in this category [97], which can discriminate acetyl and trimethyl lysine histone modifications. In summary, mass spectrometry constitutes a very powerful approach for the genome-wide profiling of histone modifications, but there is still a need for the

development of more accurate bioinformatics approaches to allow a more comprehensive and thorough study of MS histone modification spectra.

25.3 Bioinformatics Approaches for Analyzing Genome-Wide Methylation Profiling

DNA methylation, which is the only epigenetic phenomena involving the direct modification of genome underlying structure, can be profiled experimentally with bisulfite sequencing [98, 99], bisulfite microarrays [100, 101], and enrichment methods, such as MeDIP-seq and MethylCap-seq [102–104]. Different computational approaches have been developed for processing genome-wide profiling data obtained with each of the abovementioned techniques. In the case of bisulfite sequencing data, methylated cytosines are protected from chemical modification – i.e., sulfonation – induced by treatment with bisulfite, while unmethylated cytosines are sulfonated and appear as thymines after sequencing. Following, the reads obtained at the sequencing stage are mapped back to the reference genome, and the ratios of Cs and Ts are measured, representing the methylation levels of genomic regions. In principle, aligners such as those currently used for mapping ChIP-seq reads (please see in the previous section in this chapter) can be used for processing bisulfite sequencing reads, but in this case it is necessary to account for the underrepresentation of unmethylated Cs. Moreover, different approaches specifically suited for analyzing this data have been developed, comprising RRBSMAP [105], RMAP [106], GSNAP [107], and Segemehl [108], among others, which have been coined as wildcard aligners. These tools offer multiple functionalities for wildcarding Cs in the sequencing reads during the alignment and also adjusting the matrices used for scoring tag alignment for accommodating base mismatches. Furthermore, wildcard aligners allow the efficient and fast alignment to large genomic regions, although they tend to overestimate highly methylated regions. A second group of tools (MethylCoder [109], BRAT [110], and Bismark [111]) follow a more straightforward strategy, leveraging from well-established short-read alignment tools, and use a three-letter alphabet – i.e., considering T, G, and A – in the alignment. Three-letter alignment approaches are not very efficient for scanning large genomic regions, as a significant proportion of regions are filtered out of the alignment due to lack of sequence complementarity, caused by an increased alignment ambiguity. Once bisulfite sequence reads are aligned to the reference genome, the methylation levels of specific genomic regions can be estimated by using variant caller algorithms, which allow the quantitation of the frequency of Cs and Ts. For instance, Bis-SNP [112] relies on a Bayesian inference approach to evaluate strand-specific base calls and base call quality scores, and experiment-specific bisulfite conversion efficiency to derive fairly accurate DNA methylation estimates. Faster variant callers have been developed, including MethylExtract [113] that implements a modified version of the VarScan algorithm [114], and BS-SNPer [115] based on a “dynamic matrix algorithm” and Bayesian modeling, which are able to process large quantities of genomic sequences.

The most widely used bisulfite microarrays are Illumina® Infinium Methylation Assay [100], which allows single-CpG-site resolution quantitative measurement of genome-wide methylation profiles. In this assay, cytosine methylation at CpG islands is detected by multiplexed genotyping of bisulfite-converted genomic DNA, upon treatment with bisulfite (this technique also relies on bisulfite selective DNA modification of unmethylated regions, as described above). The assay uses two site-specific probes, one for methylated and another for the unmethylated *loci*. The Infinium MethylationEPIC BeadChip Kit enables quantitative genome-wide profiling of almost 900,000 methylation sites at the single-nucleotide resolution, encompassing expert-selected coverage of up to 99% of RefSeq genes, 95% of CpG islands, and ENCODE enhancer regions. In addition to the great potential of this technology, it has been the focus of intense research for the development of proprietary and open-source bioinformatics tools for processing Illumina Methylation Arrays. The GenomeStudio software developed by the chip supplier enables differential methylation analysis for small-scale studies, also including advanced tools for visualization of large amounts of data, plotting, and statistical analysis. The R/Bioconductor BeadArray toolkit [116] is also available for performing large-scale stand-alone analysis requiring more intense calculations or parallel computing infrastructures. Infinium® arrays include multiple probes for performing sample-dependent and sample-independent data quality control, which is the input of packages like IMA [117] and LumiWCluster [118]. These tools use different approaches for removing noisy probes from the chip data, which are straightforwardly filtered out based on the median detection p-value cutoff in the case of IMA, while LumiWCluster relies on a more sophisticated weighted likelihood model based on clustering methylation data. Background correction should also be performed for removing nonspecific signals and differences between replicates. This step can be performed with the GenomeStudio Infinium integrated package, but also with many other toolkits, such as lumi [119], limma [120], and BeadArray [116]. After the initial quality control, microarray data need to be normalized to remove random noise, technical artifacts, and measurement variation inherent to microarrays. Normalization should be performed between different replicate array measurements, i.e., between array, and internally for each array, i.e., within array. This can be accomplished with HumMethQCReport [121] and lumi [119], which use spline and weighted scatter smoothing for normalizing methylation data, but there are also many other alternative approaches based on different statistical approaches [122]. Special interest should also be put on scaling the signal obtained for the two different probes used in this technique – i.e., probes for methylated and unmethylated *loci* – that produce rather different signal distributions, due to the bias towards CpG islands in the genome [100]. Peak rescaling is usually performed with methods such as SWAN [123] that implements a sub-quantile within-array normalization (SQN) procedure, similar to the rationale followed in another study implementing a pipeline for processing Illumina® Infinium Methylation BeadChip [124]. Other approaches use variations of this procedure, such as the mixture quantile normalization method to rescale the distributions of the methylation and unmethylation probes into distributions that can be compared statistically [125, 126]. Batch effects, which

are also common on DNA methylation arrays, can be corrected with toolkits like CpGassoc [127], MethLAB [128], and ISVA [129] R/Bioconductor packages.

Enrichment techniques, such as MeDIP-seq and MethylCap-seq [102–104], are based on the use of proteins that specifically bind to methylated DNA regions – e.g., 5-methylcytosine-specific antibodies [104, 130] (methylated DNA immunoprecipitation (MeDIP)) or methyl-binding domain proteins [131, 132] (MethylCap) – to enrich hypermethylated fragments that are subject to HTP or microarray sequencing. The bioinformatics processing of methylation data generated with these approaches can be performed with the same methods describe above for processing sequencing or microarray platforms. Moreover, there are some methods exclusively tailored for enrichment data, like MEDIPS [133], an R/Bioconductor suite that enables processing multiple replicates and performing a great variety of statistical analyses. Another toolkit, coined as Batman [102], which stands for “Bayesian tool for methylation analysis” relies on the knowledge that almost all DNA methylation in mammals occurs at CpG dinucleotides and uses a standard Bayesian inference approach to estimate the posterior distribution of the methylation state parameters from data to generate quantitative methylation profiles. A very interesting study built on a thorough comparison of more than 20 different software tools has resulted in the development of RnBeads [134], an integrative suite that supports all genome-scale and genome-wide DNA methylation assays, implemented to facilitate stand-alone running of complex pipelines in high-performance computing infrastructures. With this toolkit, it is possible to perform all the steps of DNA methylation data analysis, ranging from data visualization, quality control, handling batch effects, correction for tissue heterogeneity, and differential DNA methylation analysis.

25.4 Computational Analysis of Chromatin Accessibility Data

The chromatin accessibility of genomic regions can be profiled with methodologies such as DNase-seq [135], FAIRE-seq [136], and ATAC-seq [137], which rely on different experimental principles and produce rather different data outputs. DNase-seq and ATAC-seq are based on the use of endonucleases – i.e., DNase I and engineered Tn5 transposase, respectively – to fragment DNA, while FAIRE-seq is a physical fragmentation method, in which DNA is treated with formaldehyde to cross-link chromatin. The differences between DNA fragmentation procedures used in each technique – i.e., DNase I and engineered Tn5 transposase have a tendency to cleave some DNA sequences more efficiently than others, and sonication could produce under and over sonicated chromatin depending on the sonication parameters used – cause that each technique generates rather different accessibility profiles [138]. In accordance, these differences should be taken into consideration while performing the downstream bioinformatics processing of sequencing data. Chromatin accessibility peaks are generally different from peak signals generated with histone modification ChIP-seq experiments, which are in general broad sequence read peaks. Hence, peak callers designed for ChIP-seq need some fine-tuning for processing chromatin accessibility data [138, 139]. Furthermore,

ChIP-seq data usually shows a higher signal-to-noise ratio compared to DNase-seq, making ChIP-seq peaks easier to detect [140]. Different peak callers have been developed to process accessibility data, including F-Seq [74] toolkit, which can be used for ChIP-seq and FAIRE-seq data [141], and ZINBA [142], which relies on a mixture regression approach for probabilistically identifying real and artifact peaks and can also handle ChIP-seq and FAIRE-seq data. Moreover, the Hotspot program [143] has been developed as part of the ENCODE project specifically for analyzing DNase-seq data, and follows a similar rationale to ChIP-seq sliding-window peak callers described above, using a probabilistic model to classify peaks by assessing the differences between the sample and a background distribution. MACS [73], which is commonly used for ChIP-seq data, and ChIPOTle [56], suited for processing ChIP-on-chip data as described above, have also been used for DNase-seq [144] and FAIRE-seq [136], respectively. In general, most of these tools have also been applied for ATAC-seq data analysis, but there are some other tools specifically implemented for this novel technique, such as I-ATAC (<https://www.jax.org/research-and-faculty/tools/i-atac>). This tool integrates multiple methods for quality check, preprocessing, and running sequential, multiple-parallel, and customized data analysis pipelines into a cross platform and open-source desktop application. Interestingly, the selection of the peak caller of use could play a key role in peak assignment output, as a comparison of the most common tools for processing accessibility data has shown that there is little overlap among called peaks obtained for the same chromatin accessibility dataset [140].

25.5 Epigenomic Databases and Epigenome Mapping Initiatives

The great developments of high-throughput sequencing technologies have allowed the steady generation of great quantities of epigenomic data in different cell types/lines and multiple organisms. This has been boosted by many large-scale epigenome mapping projects, such as the ENCODE project [145], the NIH Roadmap Epigenomics [146], the International Human Epigenome Consortium (<http://ihc-epigenomes.org/>), and the HEROIC European project (http://cordis.europa.eu/project/rcn/78439_en.html), among others. Other resources, such as the MethBase database (<http://smithlabresearch.org/software/methbase/>) [147], encompassing hundreds of methylomes from different organisms allow comparing the methylation profiles of genomics regions in different animal and plant genomes. There exist other more specialized epigenomic projects and databases encompassing information of the brain. These neuroepigenomic resources include MethylomeDB database (<http://www.neuroepigenomics.org/methylomedb>) [148] that includes genome-wide DNA methylation profiles of human and mouse brain and is integrated with a genome browser which allows surfing through the genome and analyzes the methylation of specific *loci*, searches for specific methylation profiles, and compares methylation patterns between individual samples. The Brain Cloud (<http://braincloud.jhmi.edu/>) [149] compiles methylation data from human

postmortem dorsolateral prefrontal cortices from normal subjects across the life span, also integrating single-nucleotide polymorphism data. The great amount of data generated in these projects has prompted the development of a great variety of computational tools for the analysis of epigenetic data, some of which have been described in detail in previous sections of this chapter. Moreover, the wealth of data in these databases has enabled groundbreaking studies, such as one recent report [38] encompassing a thorough integrative study of different epigenetic phenomena – e.g., chromatin accessibility, DNA methylation, chromatin marks, gene expression – in different reference epigenomes. In this study, the authors profile cells from different tissues and organs in more than 100 adult and fetal epigenomes and were able to identify epigenetic differences arising during lineage specification and cellular differentiation, which are the modules of regulatory regions with coordinated activity across cell types, and the role of regulatory regions in human disease associated with common traits and disorders [38]. This study shows that genomic regions vary greatly in their association with active marks, with approximately 5% of each epigenome marked by enhancer or promoter signatures, showing increased association with expressed genes and increased evolutionary conservation, while two-thirds of each reference epigenome are quiescent and enriched in gene-poor stably repressed regions [38]. Furthermore, the authors find that genetic variants associated with complex traits are highly enriched in epigenomic annotations of trait-relevant tissues, and genome-wide association enrichments are significantly strongest for enhancer-associated marks, consistent with their high tissue-specific nature [38]. However, promoter-associated and transcription-associated marks were also enriched, implicating several gene-regulatory levels as underlying genetic variants associated with complex traits [38].

25.6 Epigenetic Differential Analysis and Integration of Epigenomic and Gene Expression Data

Despite the great wealth of epigenomic data, we still lack a systematic understanding of how the epigenomic landscape regulates gene expression and which are the epigenetic signatures that control the most important regulatory circuitry in the transcriptional level. Differential analysis of ChIP-seq genome-wide profiles obtained for different cellular phenotypes is a rather challenging problem, due to the significant heterogeneity in peak calling between different measurements and the lack of overlap between peak assignments obtained with different peak callers [140]. The *diffReps* program [150] has been designed to detect differential sites from ChIP-seq data, with or without biological replicates, and implements a sliding-window approach to estimate the statistical significance of differential peaks based on a binomial distribution model across samples. The differential histone modification profiles generated with *diffReps* can be used to try to superimpose the epigenetic differential profile with gene expression data. The *GeneOverlap* R/Bioconductor tool implements different statistical models for estimating the significance of the overlap of histone modification and gene expression profiles. However, the great

complexity of the histone code, and the cross talk established between different histone marks to cooperatively regulate gene expression, makes it difficult to capture the regulatory epigenetic mechanisms just by superimposing histone modification and gene expression data. More complex computational models for predicting gene expression from complex histone modification profiles have been proposed [151, 152]. In order to reproduce the quantitative relationship between gene expression levels and histone modifications, these approaches combine information from many different data tracks of repressive and activating chromatin modifications, which are processed with machine learning approaches and were able to explain a fairly high proportion of the gene expression profiles in different organisms [151, 152]. In more complex expression datasets, such as brain tissues, similar approaches for combining histone modification data [153] have not been able to obtain a good correlation with the observed gene expression profiles, which could be related to the great complexity of gene regulation in these heterogeneous tissues, and the regulatory role of other histone marks not included in the study.

The prediction of epigenetic states has also been the focus of intense research. Several computational approaches have been devised for predicting promoter regions (extensively reviewed in [154]), prediction of CpG islands [155, 156], DNA methylation [157, 158], and nucleosome positioning [159, 160]. However, with the advent of next-generation sequencing (NGS), which is used in combination with techniques for profiling chromatin accessibility, histone modifications, and DNA methylation that have allowed the generation of huge quantities of genome-wide epigenetic data, the prediction of epigenetic states has lost relevance. Nevertheless, a different group of approaches has been developed for leveraging from genome annotation data at the epigenetic level for predicting the chromatin states – e.g., poised or strong enhancers, active promoters, and heterochromatin, among others – from histone modification data [161, 162]. ChromHMM [161] relies on a multivariate hidden Markov model that represents the observed combination of chromatin marks as the product of independent Bernoulli random variables for segmenting the genome into regions with different chromatin states. Segway [162] can also input histone modification data, but also DNA methylation and chromatin accessibility data, and implements a Dynamic Bayesian Network model for hierarchical genome segmentation. Interestingly, ChromHMM and Segway can be used to process fairly complex datasets of experimental data and perform chromatin state assignments, which have provided key insights in transversal epigenomic studies in different cell types, tissues, or human populations [38, 163, 164].

25.7 Systems Biology Approaches and Reconstruction of Multilevel Regulatory Networks

The availability of highly detailed annotation of human and mouse genomes [38, 145, 146] has paved the way for performing studies for integrating multilevel biological data, encompassing epigenetics, DNA sequence variation, gene expression,

and clinical data. The regulatory events triggering phenotypic transitions such as cellular differentiation, and the dysfunctions associated to disease onset and progression are usually mediated by multiple genes, which establish complex interaction networks. Thus, in order to gain understanding of the regulatory mechanisms at the epigenetic and transcriptional levels involved in the regulation of these cellular phenotypes, it is necessary to derive more comprehensive systems-level computational models. For such large-scale molecular datasets, several network approaches have been developed to identify and dissect the underlying “interactomes” for discovering key mechanisms and causal regulators in normal or pathological biological systems [165]. Gene regulatory Boolean network models have been very useful for conducting systems-level modeling of complex high-throughput biological data enabling the construction of complex interaction networks for studying disease mechanisms [166]. Disease network models have been essential for predicting disease-related genes based on the analysis of different topological characteristics, such as node connectivity [167, 168], gene-gene interaction tendency in specific tissues [169], or network neighbors of disease-related genes [170, 171]. A different group of approaches tries to model cellular phenotypes as attractors in the gene expression landscape, and phenotypic transitions are modeled by identifying nodes destabilizing these attractors [172–174], and disease perturbations, such as chemical compounds or mutations, can cause a switch from a healthy to a disease attractor state [175–177]. Co-expression-based network inference approaches [178, 179] have also been used to build regulatory network models from HTS data. Weighted gene co-expression models (WGCNA) [180] – i.e., there exists a widely used and very efficient R/Bioconductor package to build WGCNA network models [181] – which allow embodying important information of the underlying relationships and interactions among genes have been widely used to identify disease-causing genes in multigene human pathologies, such as autism [182–184] and Alzheimer’s disease [185, 186]. These WGCNA formalisms allow the generation of fairly complex network representations – e.g., eigengene networks [187, 188], in which the nodes are composite network modules. WGCNA models have enabled the identification of an age-related co-methylation module present in multiple human tissues, including the blood and brain from the analysis of up to 2442 Illumina DNA methylation arrays [189]. Similarly, these approaches have been used to identify common methylation patterns correlated with age in identical twins [190], the identification of the upstream epigenetic control and the downstream cellular physiology associated with alcohol dependence and neuroadaptive changes in alcoholic brain [191] and the prediction of the co-methylation modules associated with the Huntington’s disease pathogenesis [192]. The developments of the abovementioned integrative and other multiscale network modeling approaches for trying to integrate complex and multidimensional biological data to infer regulatory relationships linking different regulatory levels – e.g., DNA sequence variations, epigenetic, transcriptional, and metabolic – will be key for gaining a deeper understanding of disease onset and progression, or other important biological processes, such as development.

25.8 The Advent of the Single-Cell Era in Neuroepigenetics: Challenges for Analyzing Single-Cell Epigenomic Data

The great technological advances in the methodologies for generating high-quality genome-wide epigenomic data have caused a revolution in the study of the epigenetic mechanisms regulating gene expression, stem cell differentiation, disease onset and progression, and other key biological phenomena. These developments have also contributed to the emergence of the field of “neuroepigenetics,” aimed at studying the epigenetic regulatory mechanisms in cells from the central nervous system. It has been shown that in neurons, which live throughout most of the life span of an animal, epigenetic mechanisms play a key role in the regulation of the complex metabolic and gene expression these cells must go through upon synaptic input or interactions with other nervous system cells [193, 194]. One of the main problems for studying cells from the mammalian nervous systems is trying to disentangle the great cellular heterogeneity of brain tissues [195–197]. In this regard, most of the neuroepigenomic studies conducted so far have been performed with the traditional techniques for profiling chromatin accessibility, histone modifications, and DNA methylation described in this and other chapters of this book. These approaches require as input samples containing hundreds of thousands or millions of cells, encompassing highly heterogeneous cell populations. In recent years, different experimental techniques have been developed for studying heterogeneous cell populations. Gene expression single-cell transcriptional profiling techniques first developed 20 years ago [198] have become a very popular technique conventionally used in most laboratories, thanks to great technological developments in cell capture and next-generation sequencing approaches. The application of single-cell gene transcriptomics techniques has been central in the study of gene expression and functional diversity in somatosensory neurons from the dorsal root ganglia [199, 200], in different cortical regions [197, 201, 202], and developing retina [203].

Different single-cell epigenomic approaches have been recently developed for high-throughput genome-wide mapping of DNA methylation, histone modifications, and chromatin accessibility. The single-cell reduced-representation bisulfite sequencing (scRRBS) technique [204] is highly sensitive and can detect the methylation status of up to 1.5 million CpG sites within the genome of an individual cell. This technique is very efficient for profiling promoter regions, though it has poor coverage in enhancer regions. Bisulfite single-cell sequencing approaches enable genome-wide profiling of single cells or very small cell populations, although with a rather low sequencing coverage [205, 206]. Histone modification single-cell profiling can be measured with different barcoding approaches, taking advantage of techniques for indexing regions bearing the posttranslational modification in individual cells with specific sequence tags, and then performing ChIP-seq measurement after pooling cells from different wells – i.e., the heterogeneous population – which reduces the problem associated to input sample requirement of ChIP-seq [207, 208]. A different technique has been developed (the nano-ChIP-seq protocol) [209], which combines a high-sensitivity small-scale ChIP assay tailored for HTS libraries from scarce amounts of ChIP DNA. Recently, the single-tube

DNA amplification method (LinDA) has been conceived, enabling ChIP-seq measurements of picogram DNA amounts obtained from a few thousand cells [210]. Chromatin accessibility single-cell profiling can be performed with a modification of the ATAC-seq approach, based on combinatorial indexing for barcoding populations of nuclei in different wells, and then performing chromatin accessibility after pooling [211]. There exists another methodology available for single-cell chromatin accessibility profiling, based on a programmable microfluidics platform for capturing and analyzing cells in specific microfluidic chambers [212]. These methodologies are still under development for improving single-cell isolation [203, 213] and single-molecule sequencing techniques [214, 215], to try to increase the reliability of the measurements and sequencing coverage. The application of these approaches to study central nervous system samples will be essential for obtaining a clearer picture of the epigenetic regulatory mechanisms in neurons from different brain regions and how the heterogeneity at the epigenetic level defines different circuitries at the transcriptional regulatory level in central nervous system cells. However, the computational analysis of single-cell epigenomic data poses many computational challenges that will be the focus of intense research in the next years to match the great developments of experimental techniques. Currently, the computational tools and approaches used for processing single-cell epigenomic data are essentially those developed for bulk measurements, which have been thoroughly discussed in this chapter. Nevertheless, it is crucial to develop computational methods that are tailored specifically for processing single-cell data for tackling the problems associated with normalization and cell-type identification and for dissecting variability levels across cells [216]. It is expected that such methods will be developed in the next few years, leading to new discoveries in areas ranging from the physiology of tissues to systems biology [216].

References

1. Gehring M, Reik W, Henikoff S. DNA demethylation by DNA repair. *Trends Genet.* 2009;25(2):82–90.
2. Loizou JI, Murr R, Finkbeiner MG, Sawan C, Wang ZQ, Herceg Z. Epigenetic information in chromatin: the code of entry for DNA repair. *Cell Cycle.* 2006;5(7):696–701.
3. Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol.* 2009;10(3):192–206.
4. Richards EJ. Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev Genet.* 2006;7(5):395–401.
5. Grewal SI, Moazed D. Heterochromatin and epigenetic control of gene expression. *Science.* 2003;301(5634):798–802.
6. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet.* 2003;33(Suppl):245–54.
7. Heard E, Martienssen RA. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell.* 2014;157(1):95–109.
8. Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature.* 2012;484(7394):339–44.
9. Kouzarides T. Chromatin modifications and their function. *Cell.* 2007;128(4):693–705.
10. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002;16(1):6–21.

11. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012;13(7):484–92.
12. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet.* 2013;14(3):204–20.
13. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature.* 2000;403(6765):41–5.
14. Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol.* 2014;15(11):703–8.
15. Rando OJ. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr Opin Genet Dev.* 2012;22(2):148–55.
16. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol.* 2015;16(9):519–32.
17. Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature.* 2015;517(7536):640–4.
18. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature.* 2007;448(7154):714–7.
19. 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.
20. Musselman CA, Lalonde ME, Cote J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol.* 2012;19(12):1218–27.
21. Torres IO, Fujimori DG. Functional coupling between writers, erasers and readers of histone and DNA methylation. *Curr Opin Struct Biol.* 2015;35:68–75.
22. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res.* 2011;21(3):381–95.
23. Handel AE, Ebers GC, Ramagopalan SV. Epigenetics: molecular mechanisms and implications for disease. *Trends Mol Med.* 2010;16(1):7–16.
24. Ordovas JM, Smith CE. Epigenetics and cardiovascular disease. *Nat Rev Cardiol.* 2010;7(9):510–9.
25. Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol.* 2010;28(10):1057–68.
26. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* 2007;447(7143):433–40.
27. Haas J, Frese KS, Park YJ, Keller A, Vogel B, Lindroth AM, et al. Alterations in cardiac DNA methylation in human dilated cardiomyopathy. *EMBO Mol Med.* 2013;5(3):413–29.
28. Pujadas E, Feinberg AP. Regulated noise in the epigenetic landscape of development and disease. *Cell.* 2012;148(6):1123–31.
29. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet.* 2006;7(1):21–33.
30. Ahuja N, Easwaran H, Baylin SB. Harnessing the potential of epigenetic therapy to target solid tumors. *J Clin Invest.* 2014;124(1):56–63.
31. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer.* 2011;11(10):726–34.
32. Jones PA, Baylin SB. The epigenomics of cancer. *Cell.* 2007;128(4):683–92.
33. Jakovcevski M, Akbarian S. Epigenetic mechanisms in neurological disease. *Nat Med.* 2012;18(8):1194–204.
34. Day JJ, Sweatt JD. Epigenetic mechanisms in cognition. *Neuron.* 2011;70(5):813–29.
35. Dulac C. Brain function and chromatin plasticity. *Nature.* 2010;465(7299):728–35.
36. Jensen LR, Amende M, Gurok U, Moser B, Gimmel V, Tzschach A, et al. Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am J Hum Genet.* 2005;76(2):227–36.
37. Miller CA, Gavin CF, White JA, Parrish RR, Honasoge A, Yancey CR, et al. Cortical DNA methylation maintains remote memory. *Nat Neurosci.* 2010;13(6):664.
38. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539):317–30.

39. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129(4):823–37.
40. Buck MJ, Lieb JD. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*. 2004;83(3):349–60.
41. Huebert DJ, Kamal M, O'Donovan A, Bernstein BE. Genome-wide analysis of histone modifications by ChIP-on-chip. *Methods*. 2006;40(4):365–9.
42. Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*. 2012;13(12):840–52.
43. Nakato R, Shirahige K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Briefings Bioinform*. 2016.
44. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009;10(10):669–80.
45. Bartke T, Borgel J, DiMaggio PA. Proteomics in epigenetics: new perspectives for cancer research. *Brief Funct Genomics*. 2013;12(3):205–18.
46. Garcia BA, Shabanowitz J, Hunt DF. Characterization of histones and their post-translational modifications by mass spectrometry. *Curr Opin Chem Biol*. 2007;11(1):66–73.
47. Moradian A, Kalli A, Sweredoski MJ, Hess S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics*. 2014;14(4–5):489–97.
48. Tian Z, Tolic N, Zhao R, Moore RJ, Hengel SM, Robinson EW, et al. Enhanced top-down characterization of histone post-translational modifications. *Genome Biol*. 2012;13(10):R86.
49. Olsen JV, Mann M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics*. 2013;12(12):3444–52.
50. Bock C, Lengauer T. Computational epigenetics. *Bioinformatics*. 2008;24(1):1–10.
51. Zhang ZD, Rozowsky J, Lam HY, Du J, Snyder M, Gerstein M. Telescope: online analysis pipeline for high-density tiling microarray data. *Genome Biol*. 2007;8(5):R81.
52. Ji H, Wong WH. TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics*. 2005;21(18):3629–36.
53. Toedling J, Skylar O, Krueger T, Fischer JJ, Sperling S, Huber W. Ringo--an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinform*. 2007;8:221.
54. Scacheri PC, Crawford GE, Davis S. Statistics for ChIP-chip and DNase hypersensitivity experiments on NimbleGen arrays. *Methods Enzymol*. 2006;411:270–82.
55. Keles S. Mixture modeling for genome-wide localization of transcription factors. *Biometrics*. 2007;63(1):10–21.
56. Buck MJ, Nobel AB, Lieb JD. ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. *Genome Biol*. 2005;6(11):R97.
57. Li W, Meyer CA, Liu XS. A hidden Markov model for analyzing ChIP-chip experiments on genome tiling arrays and its application to p53 binding sequences. *Bioinformatics*. 2005;21(Suppl 1):i274–82.
58. Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, Brown M, et al. Model-based analysis of tiling-arrays for ChIP-chip. *Proc Natl Acad Sci U S A*. 2006;103(33):12457–62.
59. 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.
60. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
61. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*. 2009;25(15):1966–7.
62. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res*. 2002;12(4):656–64.
63. Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform*. 2010;11(5):473–83.
64. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007;448(7153):553–60.
65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.

66. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglu S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;22(9):1813–31.
67. Hansen P, Hecht J, Ibrahim DM, Krannich A, Truss M, Robinson PN. Saturation analysis of ChIP-seq data for reproducible identification of binding peaks. *Genome Res.* 2015;25(9):1391–400.
68. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. *Nat Methods.* 2009;6(11 Suppl):S22–32.
69. Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol.* 2008;26(11):1293–300.
70. 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.
71. Bardet AF, Steinmann J, Bafna S, Knoblich JA, Zeitlinger J, Stark A. Identification of transcription factor binding sites from ChIP-seq data at high resolution. *Bioinformatics.* 2013;29(21):2705–13.
72. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol.* 2008;26(12):1351–9.
73. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137.
74. Boyle AP, Guinney J, Crawford GE, Furey TS. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics.* 2008;24(21):2537–8.
75. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglu S, et al. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods.* 2008;5(9):829–34.
76. Zhang X, Robertson G, Krzywinski M, Ning K, Droit A, Jones S, et al. PICS: probabilistic inference for ChIP-seq. *Biometrics.* 2011;67(1):151–63.
77. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, 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.
78. Ibrahim MM, Lacadie SA, Ohler U. JAMM: a peak finder for joint analysis of NGS replicates. *Bioinformatics.* 2015;31(1):48–55.
79. Zhang Y, Lin YH, Johnson TD, Rozek LS, Sartor MA. PePr: a peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data. *Bioinformatics.* 2014;30(18):2568–75.
80. Tuteja G, White P, Schug J, Kaestner KH. Extracting transcription factor targets from ChIP-Seq data. *Nucleic Acids Res.* 2009;37(17):e113.
81. Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, et al. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol.* 2009;27(1):66–75.
82. Flensburg C, Kinkel SA, Keniry A, Blewitt ME, Oshlack A. A comparison of control samples for ChIP-seq of histone modifications. *Frontiers in genetics.* 2014;5:329.
83. Laajala TD, Raghav S, Tuomela S, Lahesmaa R, Aittokallio T, Elo LL. A practical comparison of methods for detecting transcription factor binding sites in ChIP-seq experiments. *BMC Genomics.* 2009;10:618.
84. Malone BM, Tan F, Bridges SM, Peng Z. Comparison of four ChIP-Seq analytical algorithms using rice endosperm H3K27 trimethylation profiling data. *PLoS One.* 2011;6(9):e25260.
85. Horn DM, Zubarev RA, McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J Am Soc Mass Spectrom.* 2000;11(4):320–32.
86. Liu X, Inbar Y, Dorrestein PC, Wynne C, Edwards N, Souda P, et al. Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol Cell Proteomics.* 2010;9(12):2772–82.
87. Carvalho PC, Xu T, Han X, Cociorva D, Barbosa VC, Yates 3rd. JR. YADA: a tool for taking the most out of high-resolution spectra. *Bioinformatics.* 2009;25(20):2734–6.

88. Huttenhain R, Hess S. A combined top-down and bottom-up MS approach for the characterization of hemoglobin variants in Rhesus monkeys. *Proteomics*. 2010;10(20):3657–68.
89. Sidoli S, Cheng L, Jensen ON. Proteomics in chromatin biology and epigenetics: Elucidation of post-translational modifications of histone proteins by mass spectrometry. *J Proteomics*. 2012;75(12):3419–33.
90. Zamborg L, LeDuc RD, Glowacz KJ, Kim YB, Viswanathan V, Spaulding IT, et al. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic acids research*. 2007;35(Web Server issue):W701–6.
91. Liu X, Sirotkin Y, Shen Y, Anderson G, Tsai YS, Ting YS, et al. Protein identification using top-down. *Mol Cell Proteom*. 2012;11(6):M111 008524.
92. Kalli A, Sweredoski MJ, Hess S. Data-dependent middle-down nano-liquid chromatography-electron capture dissociation-tandem mass spectrometry: an application for the analysis of unfractionated histones. *Anal Chem*. 2013;85(7):3501–7.
93. DiMaggio Jr PA, Young NL, Baliban RC, Garcia BA, Floudas CA. A mixed integer linear optimization framework for the identification and quantification of targeted post-translational modifications of highly modified proteins using multiplexed electron transfer dissociation tandem mass spectrometry. *Mol Cell Proteomics*. 2009;8(11):2527–43.
94. Pesavento JJ, Mizzen CA, Kelleher NL. Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: human histone H4. *Anal Chem*. 2006;78(13):4271–80.
95. Siuti N, Roth MJ, Mizzen CA, Kelleher NL, Pesavento JJ. Gene-specific characterization of human histone H2B by electron capture dissociation. *J Proteome Res*. 2006;5(2):233–9.
96. Guan S, Burlingame AL. Data processing algorithms for analysis of high resolution MSMS spectra of peptides with complex patterns of posttranslational modifications. *Mol Cell Proteomics*. 2010;9(5):804–10.
97. Matthiesen R, Trelle MB, Hojrup P, Bunkenborg J, Jensen ON. VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J Proteome Res*. 2005;4(6):2338–47.
98. Booth MJ, Branco MR, Ficiz G, Oxley D, Krueger F, Reik W, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*. 2012;336(6083):934–7.
99. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol*. 2009;27(4):353–60.
100. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98(4):288–95.
101. Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, et al. Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res*. 2006;34(2):528–42.
102. Down TA, Rakyanc VK, Turner DJ, Flicek P, Li H, Kulesha E, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol*. 2008;26(7):779–85.
103. Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res*. 2010;38(2):391–9.
104. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet*. 2005;37(8):853–62.
105. Xi Y, Bock C, Muller F, Sun D, Meissner A, Li W. RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing. *Bioinformatics*. 2012;28(3):430–2.
106. Smith AD, Chung WY, Hodges E, Kendall J, Hannon G, Hicks J, et al. Updates to the RMAP short-read mapping software. *Bioinformatics*. 2009;25(21):2841–2.

107. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*. 2010;26(7):873–81.
108. Otto C, Stadler PF, Hoffmann S. Fast and sensitive mapping of bisulfite-treated sequencing data. *Bioinformatics*. 2012;28(13):1698–704.
109. Pedersen B, Hsieh TF, Ibarra C, Fischer RL. MethylCoder: software pipeline for bisulfite-treated sequences. *Bioinformatics*. 2011;27(17):2435–6.
110. Harris EY, Ponts N, Le Roch KG, Lonardi S. BRAT-BW: efficient and accurate mapping of bisulfite-treated reads. *Bioinformatics*. 2012;28(13):1795–6.
111. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27(11):1571–2.
112. Liu Y, Siegmund KD, Laird PW, Berman BP. Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biol*. 2012;13(7):R61.
113. Barturen G, Rueda A, Oliver JL, Hackenberg M. MethylExtract: High-Quality methylation maps and SNV calling from whole genome bisulfite sequencing data. *F1000Research*. 2013;2:217.
114. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009;25(17):2283–5.
115. Gao S, Zou D, Mao L, Liu H, Song P, Chen Y, et al. BS-SNPer: SNP calling in bisulfite-seq data. *Bioinformatics*. 2015;31(24):4006–8.
116. Dunning MJ, Smith ML, Ritchie ME, Tavaré S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics*. 2007;23(16):2183–4.
117. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, et al. IMA: an R package for high-throughput analysis of Illumina's 450 K Infinium methylation data. *Bioinformatics*. 2012;28(5):729–30.
118. Kuan PF, Wang S, Zhou X, Chu H. A statistical framework for Illumina DNA methylation arrays. *Bioinformatics*. 2010;26(22):2849–55.
119. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24(13):1547–8.
120. Wettenhall JM, Smyth GK. limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics*. 2004;20(18):3705–6.
121. Mancuso FM, Montfort M, Carreras A, Alibes A, Roma G. HumMeth27QCReport: an R package for quality control and primary analysis of Illumina Infinium methylation data. *BMC Res Notes*. 2011;4:546.
122. Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerstrom-Billai F, Jagodic M, et al. An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics*. 2013;8(3):333–46.
123. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol*. 2012;13(6):R44.
124. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450 K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 2012;4(3):325–41.
125. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450 K methylation array data. *BMC Genomics* 2013;14:293.
126. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29(2):189–96.
127. Barfield RT, Kilaru V, Smith AK, Conneely KN. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics*. 2012;28(9):1280–1.
128. Kilaru V, Barfield RT, Schroeder JW, Smith AK, Conneely KN. MethLAB: a graphical user interface package for the analysis of array-based DNA methylation data. *Epigenetics*. 2012;7(3):225–9.
129. Teschendorff AE, Zhuang J, Widschwendter M. Independent surrogate variable analysis to deconvolve confounding factors in large-scale microarray profiling studies. *Bioinformatics*. 2011;27(11):1496–505.

130. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell*. 2006;126(6):1189–201.
131. Cross SH, Charlton JA, Nan X, Bird AP. Purification of CpG islands using a methylated DNA binding column. *Nat Genet*. 1994;6(3):236–44.
132. Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, et al. Transient cyclical methylation of promoter DNA. *Nature*. 2008;452(7183):112–5.
133. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics*. 2014;30(2):284–6.
134. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods*. 2014;11(11):1138–40.
135. Song L, Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor protocols*. 2010;2010(2):pdb prot5384.
136. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*. 2007;17(6):877–85.
137. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8.
138. Meyer CA, Liu XS. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat Rev Genet*. 2014;15(11):709–21.
139. Madrigal P, Krajewski P. Current bioinformatic approaches to identify DNase I hypersensitive sites and genomic footprints from DNase-seq data. *Front Genet*. 2012;3:230.
140. Koohy H, Down TA, Spivakov M, Hubbard T. A comparison of peak callers used for DNase-Seq data. *PLoS One*. 2014;9(5):e96303.
141. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets. *Nat Genet*. 2010;42(3):255–9.
142. 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.
143. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012;489(7414):75–82.
144. Wang YM, Zhou P, Wang LY, Li ZH, Zhang YN, Zhang YX. Correlation between DNase I hypersensitive site distribution and gene expression in HeLa S3 cells. *PLoS One*. 2012;7(8):e42414.
145. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
146. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28(10):1045–8.
147. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One*. 2013;8(12):e81148.
148. Xin Y, Chanrion B, O'Donnell AH, Milekic M, Costa R, Ge Y, et al. MethylomeDB: a database of DNA methylation profiles of the brain. *Nucleic Acids Res*. 2012;40(Database issue):D1245–9.
149. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011;478(7370):519–23.
150. Shen L, Shao NY, Liu X, Maze I, Feng J, Nestler EJ. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS one*. 2013;8(6):e65598.
151. Cheng C, Yan KK, Yip KY, Rozowsky J, Alexander R, Shou C, et al. A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets. *Genome Biol*. 2011;12(2):R15.

152. Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, et al. Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.* 2012; 13(9):R53.
153. Feng J, Wilkinson M, Liu X, Purushothaman I, Ferguson D, Vialou V, et al. Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol.* 2014;15(4):R65.
154. Bajic VB, Tan SL, Suzuki Y, Sugano S. Promoter prediction analysis on the whole human genome. *Nat Biotechnol.* 2004;22(11):1467–73.
155. Bock C, Walter J, Paulsen M, Lengauer T. CpG island mapping by epigenome prediction. *PLoS Comput Biol.* 2007;3(6):e110.
156. Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM. Predicting aberrant CpG island methylation. *Proc Natl Acad Sci U S A.* 2003;100(21):12253–8.
157. Das R, Dimitrova N, Xuan Z, Rollins RA, Haghghi F, Edwards JR, et al. Computational prediction of methylation status in human genomic sequences. *Proc Natl Acad Sci U S A.* 2006;103(28):10713–6.
158. Fang F, Fan S, Zhang X, Zhang MQ. Predicting methylation status of CpG islands in the human brain. *Bioinformatics.* 2006;22(18):2204–9.
159. Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K, et al. Nucleosome positioning signals in genomic DNA. *Genome Res.* 2007;17(8):1170–7.
160. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, et al. A genomic code for nucleosome positioning. *Nature.* 2006;442(7104):772–8.
161. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods.* 2012;9(3):215–6.
162. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nat Methods.* 2012;9(5):473–6.
163. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;507(7493):455–61.
164. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014; 46(3):310–5.
165. Maze I, Shen L, Zhang B, Garcia BA, Shao N, Mitchell A, et al. Analytical tools and current challenges in the modern era of neuroepigenomics. *Nat Neurosci.* 2014;17(11):1476–90.
166. Schadt EE, Friend SH, Shaywitz DA. A network view of disease and compound screening. *Nat Rev Drug Discov.* 2009;8(4):286–95.
167. Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. *Bioinformatics.* 2006;22(18):2291–7.
168. Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics.* 2005;21(23): 4205–8.
169. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci U S A.* 2007;104(21):8685–90.
170. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011;12(1):56–68.
171. Lage K, Karlberg EO, Stirling ZM, Olason PI, Pedersen AG, Rigina O, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol.* 2007;25(3):309–16.
172. Crespo I, Del Sol A. A general strategy for cellular reprogramming: the importance of transcription factor cross-repression. *Stem Cells.* 2013;31(10):2127–35.
173. Crespo I, Perumal TM, Jurkowski W, del Sol A. Detecting cellular reprogramming determinants by differential stability analysis of gene regulatory networks. *BMC Syst Biol.* 2013;7:140.
174. Del Sol A, Buckley NJ. Concise review: a population shift view of cellular reprogramming. *Stem Cells.* 2014;32(6):1367–72.
175. del Sol A, Balling R, Hood L, Galas D. Diseases as network perturbations. *Curr Opin Biotechnol.* 2010;21(4):566–71.

176. Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev Biol.* 2009;20(7):869–76.
177. Zickert S, Angarica VE, Upadhyaya BB, del Sol A. Prediction of disease-gene-drug relationships following a differential network analysis. *Cell Death Dis.* 2016;7:e2040.
178. Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. *Nat Methods.* 2012;9(8):796–804.
179. Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G. Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci U S A.* 2010;107(14):6286–91.
180. Horvath S. *Weighted Network Analysis: Applications in Genomics and Systems Biology*; Springer; 2011.
181. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics.* 2008;9:559.
182. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell.* 2013;155(5):1008–21.
183. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, et al. Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. *Am J Hum Genet.* 2012;91(1):38–55.
184. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature.* 2011;474(7351):380–4.
185. Miller JA, Horvath S, Geschwind DH. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc Natl Acad Sci U S A.* 2010;107(28):12698–703.
186. Miller JA, Woltjer RL, Goodenbour JM, Horvath S, Geschwind DH. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med.* 2013;5(5):48.
187. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol.* 2007;1:54.
188. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, et al. Functional organization of the transcriptome in human brain. *Nat Neurosci.* 2008;11(11):1271–82.
189. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, et al. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 2012;13(10):R97.
190. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, et al. Epigenetic predictor of age. *PLoS One.* 2011;6(6):e14821.
191. Ponomarev I, Wang S, Zhang L, Harris RA, Mayfield RD. Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. *J Neurosci.* 2012;32(5):1884–97.
192. Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF, et al. Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging.* 2016;8(7):1485–512.
193. Gage FH, Temple S. Neural stem cells: generating and regenerating the brain. *Neuron.* 2013;80(3):588–601.
194. Satterlee JS, Beckel-Mitchener A, Little R, Procaccini D, Rutter JL, Lossie AC. Neuroepigenomics: Resources, Obstacles, and Opportunities. *Neuroepigenetics.* 2015;1:2–13.
195. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol.* 2014;32(10):1053–8.
196. Poulin JF, Tasic B, Hjerling-Leffler J, Trimarchi JM, Awatramani R. Disentangling neural cell diversity using single-cell transcriptomics. *Nat Neurosci.* 2016;19(9):1131–41.
197. Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci.* 2016;19(2):335–46.
198. Brady G, Barbara M, Iscove NN. Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. *Methods Mol Cell Biol.* 1990;2(1):17–25.

199. Li CL, Li KC, Wu D, Chen Y, Luo H, Zhao JR, et al. Somatosensory neuron types identified by high-coverage single-cell RNA-sequencing and functional heterogeneity. *Cell Res.* 2016;26(8):967.
200. Usoskin D, Furlan A, Islam S, Abdo H, Lonnerberg P, Lou D, et al. Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci.* 2015;18(1):145–53.
201. Fuzik J, Zeisel A, Mate Z, Calvigioni D, Yanagawa Y, Szabo G, et al. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat Biotechnol.* 2016;34(2):175–83.
202. Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science.* 2015;347(6226):1138–42.
203. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell.* 2015;161(5):1202–14.
204. 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.
205. Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schonegger A, Klughammer J, et al. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* 2015;10(8):1386–97.
206. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods.* 2014;11(8):817–20.
207. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretzky I, Jaitin DA, David E, et al. Immunogenetics. Chromatin state dynamics during blood formation. *Science.* 2014;345(6199):943–9.
208. Rotem A, Ram O, Shoshani N, Sperling RA, Goren A, Weitz DA, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol.* 2015;33(11):1165–72.
209. Adli M, Bernstein BE. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nat Protoc.* 2011;6(10):1656–68.
210. Shankaranarayanan P, Mendoza-Parra MA, Walia M, Wang L, Li N, Trindade LM, et al. Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. *Nat Methods.* 2011;8(7):565–7.
211. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science.* 2015;348(6237):910–4.
212. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature.* 2015;523(7561):486–90.
213. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell.* 2015;161(5):1187–201.
214. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science.* 2009;323(5910):133–8.
215. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods.* 2010;7(6):461–5.
216. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet.* 2015;16(3):133–45.