Quantitative Biology 2022, 10(1): 55–66
https://doi.org/10.15302/J-QB-022-0289

REVIEW

The progress on the estimation of DNA methylation level and the detection of abnormal methylation

Shicai Fan[1,2,*], Likun Wang[3], Liang Liang[4], Xiaohong Cao[5], Jianxiong Tang[2], Qi Tian[2]

[1] Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518110, China
[2] School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[3] Institute of Systems Biomedicine, Beijing Key Laboratory of Tumor Systems Biology, Department of Pathology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China
[4] Cancer Center, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 611731, China
[5] Department of Geriatric Endocrinology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 611731, China
* Correspondence: shicaifan@uestc.edu.cn

Received August 12, 2021; Revised November 5, 2021; Accepted November 20, 2021

***Background*: DNA methylation is a key heritable epigenetic modification that plays a crucial role in transcriptional regulation and therefore a broad range of biological processes. The complex patterns of DNA methylation highlight the significance of the profiling the DNA methylation landscape.**
***Results*: In this review, the main high-throughput detection technologies are summarized, and then the three trends of computational estimation of DNA methylation levels were analyzed, especially the expanding of the methylation data with lower coverage. Furthermore, the detection methods of differential methylation patterns for sequencing and array data were presented.**
***Conclusions*: More and more research indicated the great importance of DNA methylation changes across different diseases, such as cancers. Although a lot of enormous progress has been made in understanding the role of DNA methylation, only few methylated genes or functional elements serve as clinically relevant cancer biomarkers. The bottleneck in DNA methylation advances has shifted from data generation to data analysis. Therefore, it is meaningful to develop machine learning models for computational estimation of methylation profiling and identify the potential biomarkers.**

Keywords: DNA methylation; genome-wide profiling; computational estimation; single-cell methylome; differential methylation detection

Author summary: With the development of experimental profiling approach for both pooled and single cells, the research on computational methods for methylome analysis is also a hot field for the understanding of epigenomic code. The computational estimation of DNA methylation levels, especially, the expanding methods for methylome data with lower coverage was intensively analyzed. With the broader range of DNA methylation landscapes both in coverage and sample size, it provides better opportunity for the identification of the potential biomarkers.

INTRODUCTION

As one of the most studied epigenetic modifications, DNA methylation plays an essential role in the transcriptional regulation. In vertebrates, DNA methylation involves the addition of a methyl group to the 5′ position of cytosine in CpG dinucleotides [1]. 70%–80% of the CpG dinucleotides are methylated, while CpG

islands which are rich in CpG contents are usually resistant to DNA methylation [1]. DNA methylation has been associated with various cellular processes, such as embryonic development and differentiation, silencing chromosomal domains and imprinting [2–8]. Aberrant methylation and unmethylation has been closely linked to various diseases, such as cancers [9,10]. Therefore, the genome-wide methylation detection is essential for the investigation of a broad range of diseases.

Recent progress of both experimental and computational profiling approaches for DNA methylation provides an unprecedentedly comprehensive view of the methylation landscape, and makes it possible to detect the differential methylation patterns in genome scale and in a large number of sample size [11–22]. Especially, the development of sequencing technology at single-cell levels opens the door to exciting new fields of research [23–27].

Excellent reviews concerning about the various experimental DNA methylation detection, preprocessing of both sequencing data and array data, the methylation biomarkers in different cancers are available [11, 28–33]. With the rapid development of DNA methylation, the current progress of methylation detection for both pooled cells and single cells, the computational estimation of DNA methylation and the differential methylation detection methods are urgently required.

In this review, we firstly described the high-throughput detection technologies, including the bisulphite sequencing method, Illumina chiparray method, long-read sequencing method and the single cell bisulfite based techniques. Then we analyzed the three directions in computational methylation estimation methods, which include the methylation prediction, methylation expansion and methylation imputation. Furthermore, the detection methods of differential methylation patterns for sequencing and array data were presented. The potential research direction in the future were analyzed in the last chapter.

GENOME-WIDE PROFILING TECHNIQUES FOR DNA METHYLATION

A large variety of genome wide DNA methylation detection methods have been developed in the last few years. The developmental route could be described in the Fig.1.

Genome-wide methylation detection for pooled cells

Sequencing based methods

The combination of bisulfite free or bisulfite based pretreatment approaches and sequencing based analysis method produced different detection methods, which include enzyme digestion sequencing, affinity enrichment sequencing and bisulfite sequencing methods.

Restriction enzyme-based methods cleave the unmethylated target sequences and leave the methylated DNA intact, and the following step would then reveal the locations of the unmethylated CpG sites within the recognition sites of the enzyme utilized [34]. The widely used enzyme based sequencing method is MRE-seq [35], which makes use of the differential digestion properties of isoschizomers and neoschizomers. This method has relatively low coverage of the genome and can only analyze the specific DNA sequences.

For the affinity enrichment method, it applies either methyl-CpG-binding domain proteins or antibodies to enrich methylated DNA regions. For example, MeDIP-seq utilizes an anti-methylcytosine antibody to immunoprecipitate DNA with methylated CpG sites [28],

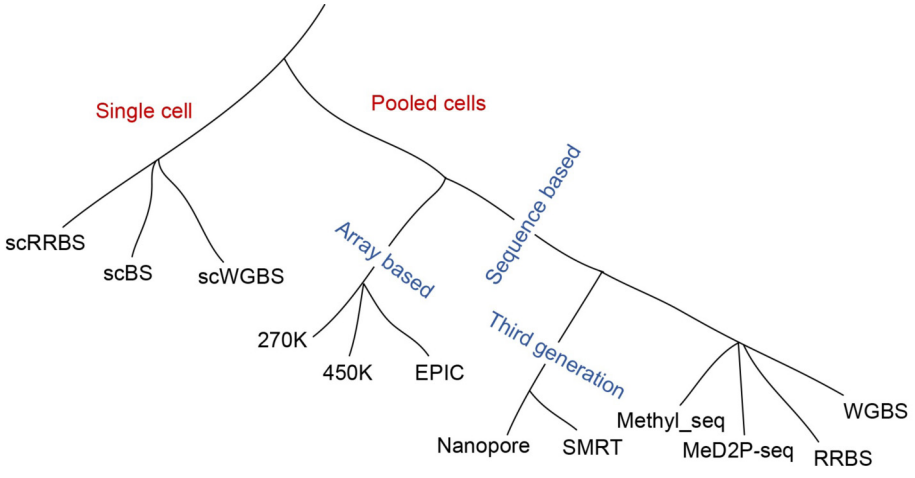

Figure 1. The development route of the genome-wide profiling techniques for DNA methylation.

which is cost-effective and feasible when there is only a low amount of starting DNA material, but it could not discriminate methylation context.

Bisulfite sequencing approach aims to investigate specific DNA sequences in single-base resolution. The treatment of genomic DNA with sodium bisulfite results in deamination of unmethylated cytosines to uracil, and leaves the methylated cytosines unaffected [28]. Whole-genome bisulfite sequencing (WGBS) theoretically covers all the cytosine information, including low CpG density regions, such as the intergenic gene deserts and distal regulatory elements. It provides accurate binary calls of cytosine methylation status, and is considered as the golden standard method in the DNA methylation study, therefore, it has been widely used in the NIH Roadmap [36] and ENCODE projects [37]. WGBS is the most comprehensive one among the existing detection methods, while there are also two limits for the WGBS method that a considerable amount of DNA is required, and the data analysis is relatively cost and difficult. To investigate the methylome at a lower cost, the reduced representation bisulfite sequencing (RRBS) method was developed. It only sequence a fraction of the genome, which integrates *Msp*I restriction enzyme digestion, enrichment of CPG-rich regions, bisulfite conversion and the sequencing for the analysis of specific fragments [38]. RRBS is less cost because it focuses on the enrichment of CpG-rich regions and is widely applied in methylation landscape of large-scale samples. The obvious drawback of RRBS is that it is limited to the loci containing *Msp*I cut sites, and it exhibits a lack of coverage at intergenic and distal regulatory elements that are relatively less studied.

Because of the mappability problem of short reads, long-read sequencing method was proposed, which mainly include the nanopore sequencing from Oxford Nanopore Technologies (ONT) and single molecule real-time (SMRT) sequencing from Pacific Biosciences. PacBio's SMRT sequencing allows the direct detection of base modification resulting from the addition of fluorescently labeled nucleotides into complementary DNA strands. It has the advantage of low requirement of input DNA, and could be applied to detect several types of epigenetic modifications, such as 6mA and 4mC [39]. In nanopore sequencing, it detects the variation in ionic current, which could distinguish the cytosine from the 5mC and 5hmC. There are mainly three steps for the nanopore sequencing, which includes base calling with canonical bases, anchoring the raw signal to a genomic reference and the identifying whether the base is modified. Its output is the probability that a base is modified at almost single-nucleotide resolution. However, because of the higher error rate and cost, these two third-generation sequencing methods are still limited for wild application.

Array based method

Besides the sequencing based detection methods, array based approaches are also widely used for methylation detection. For array based method, after the bisulfite conversion of genomic DNA and amplification, the converted DNA is hybridized to arrays containing predesigned probes to distinguish between methylated and unmethylated cytosines. Illumina adapts its GoldenGate BeadArray technology to interrogate DNA methylation in human genomic DNA samples, and produces three series of methylation beadchips, which includes Illumina HumanMethylation27 BeadChip (HM270K array), HumanMethylation450 BeadChip (HM450K array) and MethylationEPIC BeadChip (EPIC array). HM450K array covers more than 450,000 CpG methylation sites including 96% of the CGIs, 92% of the CGI shores and 86% of the CGI shelves [40]. The EPIC array covers more than 850,000 methylation sites which includes more than 90% of the HM450K sites plus additional CpG sites in the enhancer regions [41]. The HM450K array is the most widely used approach in the study of cancer methylome and other epigenomics, and the Cancer Genome Atlas (TCGA) consortium has mapped DNA methylation in thousands of cancer samples using array based methods [42]. The array based methods were cost-effective, but also have obvious disadvantages that the coverages of the HM270K, HM450K and EPIC array are about 0.9%, 1.5% and 2.8% of the whole genome.

Genome-wide methylation detection for single cell

The detection methods mentioned above could only provide the average measurements across bulk cell population, therefore, it is an open question that whether the methylation models based on the bulk data hold true when scrutinised on the single-cell level. To assess the methylation heterogeneity among individual cells, single-cell bisulfite based techniques are developed. The first single-cell reduced-representation bisulfite sequencing (scRRBS) was proposed by the integration of all steps to bisulfite conversion into one tube, followed by two rounds of PCR amplification and deep sequencing [43]. Because of the relatively poor coverage and the PCR-induced amplification bias about scRRBS, single-cell bisulfite sequencing (scBS-seq) was developed [23], which improved the measurement of DNA methylation at up to 48.4% of the CpG sites. In order to facilitate the

full genome wide coverage and trace the allele- or strand-specific methylation differences, the technology of single-cell whole-genome bisulfite sequencing (scWGBS) was then provided to infer the epigenomic cell-state dynamics in pluripotent and differentiating cells [44].

The single cell methylation technique was firstly applied to analyze the early mammalian development, which confirmed some previously established findings and also retrieved some truly novel findings [45,46]. Besides, the application of single cell methylation studies makes significant contributions to the understanding of cancer biology and the precise cancer treatment [44,47]. But still, there is a long journey for the satisfying coverage rate for current scWGBS detection technique.

COMPUTATIONAL ESTIMATION OF DNA METHYLATION LEVEL

The experimental detection methods provide the most important ways to characterize the genome wide DNA methylation patterns for better understanding of the regulatory mechanisms. To improve the efficiency and accuracy of the genome wide experimental data, some computational tools that map bisulfite sequencing data and align the bisulfite-converted reads were developed [48–50]. However, the whole genome bisulfite sequencing is expensive, labor intensive and subject to conversion bias, some immunoprecipitation sequencing methods are experimentally difficult or unfeasible in some contexts, array based methods are cost effective, but biased to gene regions and CpG islands. Therefore, the estimation of DNA methylation level with computational methods has attracted a lot attention in the last few decades. Generally, there are three directions in the computational works: prediction the methylation levels from one tissue/individual to other tissues/individuals, expanding the methylation landscape from lower coverage to higher coverage and the imputation of the single cell methylome data. The sketch map was shown in Fig. 2.

Prediction the DNA methylation levels

Although the tissue-specific gene expression patterns that determine cell types and functions are reported to be regulated partly by tissue-specific methylation, the locus specific methylation patterns between tissues were found to be highly consistent across individuals [51,52]. Lots of works have been implemented to develop prediction models that could predict the methylation levels of different tissues [53,54].

In 2006, we developed a prediction model called MethCGI [55], which was constructed with support vector machine based on the methylation data from human brain. Applying the nucleotide sequence contents as well as transcription factor binding sites (TFBSs), MethCGI achieved specificity of 84.65% and sensitivity of 84.32% in cross validation, and it also performed

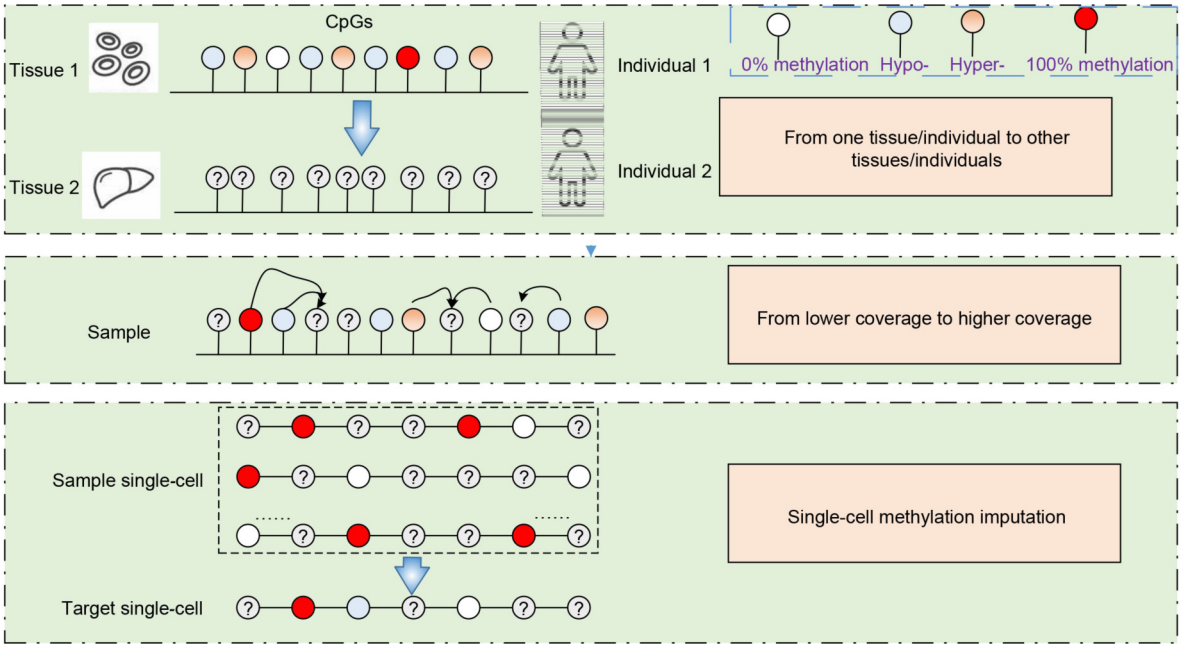

Figure 2. The three directions of computational estimation of DNA methylation level.

quite well on some other tested tissues, such as lung, liver.

Ma *et al.* developed linear prediction model and SVM prediction model on peripheral blood leukocytes (PBL), atrium and artery samples based on the methylation profiling from beadarrays [56]. The methylation values either at nearby regions or somewhere on the genome that might be correlated with a particular CpG sites were used as the modeling features. Both the cross-tissue prediction validation results and independent tissue prediction results indicated that DNA methylation values were largely conserved across tissues. Their intensive research also indicated that the prediction performance would depend on the actual tissue pairs and sample size of the training data set, and the performance could be improved if the surrogate tissue would be representative of the target tissue.

Zhang *et al.* developed a machine learning framework for prediction and characterizing DNA methylation in mammalian genomes [57]. With the minimal set of discriminative features (OFS), such as CpG statistics, repeats, histone modifications and DNase, the methylation prediction model was trained with SVM on H1 or NPC cell. With the transfer learning strategy, the SVM model was tested on tissues different from the training tissue, and the prediction results are comparable to the corresponding results in the same cell type. Also they found that the NPC trained SVM model performed well on the MSC dataset, but performed only modestly in IMR90, which indicated that there were obvious different methylation profiling between the stem cells and the cells loss of pluripotency.

To explore the performance of cross-tissue prediction intensively, the Liang group investigated the methylation concordance between placenta and cord blood [58], whose methylation correlation is lower than other tissue pairs. Using the single-CpG-based SVM model and the multiple-CpG-based model, the locus-specific DNA methylation levels in placenta were predicted using DNA methylation levels in cord blood. The results showed that a subset of CpG sites were highly correlated between measured and predicted placenta methylation levels. Therefore, it provided a reference data to predict the methylation levels on placenta which was rarely feasible to collect in large epidemiological studies.

Our group proposed a deep learning model based on convolutional neural networks for predicting DNA methylation at single-CpG-site precision [59]. In contrast to other traditional prediction tools with predefined features, our MRCNN model trained the prediction model on H1 ESC using the raw sequence as input, and applied the model to predict the genome wide methylation levels on the normal brain white matter, lung and colon tissues. The MRCNN model trained on H1 ESC data obtained high accuracy on the prediction of the methylation levels of these tissues, which indicated that MRCNN model was robust in prediction methylation levels of different types of tissues.

Expanding the DNA methylation profiling

Some methylation detection methods only provide partial methylation landscape, it is curious to know whether it is possible to expand the detected landscape to a larger landscape, *i.e.*, expand the methylation coverage based on the detected methylation profiling.

To better use the current thousands of 450K array data, our group developed the first expanding method for 450K array data in 2016 [60]. By integrating the sequence features in flanking regions and methylation values in 450K array data, several machine learning models were developed to predict the methylation levels of CpGs uncovered by 450K array. The proposed model expanded the coverage of Illumina 450K (~11 folds) and showed superior prediction accuracies. Then by integrating information derived from the similarity of local methylation pattern between tissues, the methylation information of flanking CpG sites and the methylation tendency of flanking DNA sequences, we updated the expanding model and applied it to 450K data of rheumatoid arthritis (RA), osteoarthritis and normal FLS [61]. Our model successfully expanded the coverage of CpG sites 18.5-fold and accounts for about 30% of all the CpGs in the human genome, and identified genes and pathways tightly related to RA pathogenesis. To improve the efficiency of expansion method, our group then presented an improved model called EAGLING with a more than 10 times faster speed compared to our previous methods [42]. In EAGLING, two features related to the methylation levels were used to build a logistic regression model, and the expanded methylation profiling, gene expression, and somatic mutation data were integrated to identify the potential biomarkers of 13 cancers in TCGA. The integrative analysis using the expanded methylation data is powerful in identifying critical genes/pathways that may serve as new therapeutic targets.

The number of interrogated CpGs in repetitive elements (RE) remains quite limited for current arrays, and the profiled CpGs in RE are generally sparse. Based on the methylation levels, the genetic information and RE information in specific flanking regions, Zheng *et al.* reported a random forest-based algorithm that can accurately predict genome-wide locus-specific RE methylation based on beadarray profiling [62]. The expanding model achieved satisfying performance in the validation, and achieved 2.7–3.7 times as many Alu and about 20% more LINE-1 compared with the original

HM450/EPIC. Based on the expanded RE CpG loci, it enables more comprehensive differential methylation analyses and improves power to discriminate tumor from normal tissues.

With the intensive application of EPIC array, it is also an interesting field to expand methylation levels from HM450K array to EPIC array. Applying the methylation levels of 450K array as the explanatory variable, Li *et al.* presented an ensemble method, CUE, to get the CpG loci of EPIC that not covered in HM450K array [63]. In cross-validation, the CUE ensemble model had higher prediction accuracy and lower predicted RMSE compared with individual methods, such as KNN, logistic regression and random forest, and could obtain larger number of CpG probes passed the quality control. Accurately expanded methylation values could subsequently improve the power in downstream analysis and be helpful as new methylation arrays continue to be developed.

Considering that these expanding models concentrated on improving the overall prediction accuracy for the CpG loci while neglecting whether each locus is precisely predicted, our group developed a method for constructing precise prediction models for each single CpG locus [64]. For any interested CpG locus that presented in EPIC array but not covered in 450K array, a logistic regression algorithm called PretiMeth was built based on only one DNA methylation feature that shared the most similar methylation pattern with the CpG locus to be predicted. PretiMeth outperformed other algorithms in the prediction accuracy, and kept robust across platforms and cell types, which indicated that the precise prediction models could be probably used for reference in the probe set design when the DNA methylation beadchips update.

Besides the expanding for methylation array data, Yu *et al.* developed a mixture regression model (MRM) to expand the RRBS data [65]. By integrating information of neighboring CpGs and the similarities in local methylation patterns across subjects and across multiple genomic regions, they achieved satisfying accuracy on both simulated data and real RRBS data, and recovered the methylation levels of 300K CpGs in the promoter regions of chromosome 17.

The details of the main expanding models are shown in Table 1.

Imputation for single-cell data

The development of scBS-seq and scRRBS makes it possible to uncover the heterogeneity and dynamics of DNA methylation. However, due to the small amount of DNA in the cells, they often result in very sparse coverage of genome-wide CpG (20%–40% for scBS-seq and 1%–10% for scRRBS-seq). For the quantitative analysis of the single-cell methylome data, the computational imputation of the single-cell methylomes becomes an important preprocessing step for the downstream analysis.

DeepCpG was the first genome-wide imputation

Table 1 The detail information of the main expanding models

Method	Algorithm	Features	Original coverage	Expanded profiling	Tool
Fan *et al.* [60]	SVM	a) DNA sequence b) Neighboring methylation values	450K	~11 folds of 450K	/
Fan *et al.* [61]	LR	a) DNA sequence b) Neighboring methylation values c) Methylation values from similarity tissues	450K	~18.5 folds of 450K	http://wanglab.ucsd.edu/star/LR450K/
Fan *et al.* [42]	LR	a) Neighboring methylation values b) Methylation values from similarity tissues	450K	~18.9 folds of 450K, more than 10 times faster speed	http://114.55.236.67:8013/Integrative_Analysis/home
Zheng *et al.* [62]	RF	a) RE CpG density and RE length b) Smith-Waterman (SW) score c) Number of neighboring profiled CpGs d) Genomic region of the target CpG	450K/850K	2.7–3.7 times Alu and about 20% more LINE-1	http://bioconductor.org/packages/release/bioc/html/REMP.html
Li *et al.* [63]	Ensemble model	Methylation values in 450K	450K	850K	/
Tang *et al.* [64]	LR	Methylation feature that shared the most similar methylation pattern with the CpG locus to be predicted	450K	850K	https://github.com/JxTang-bioinformatics/PretiMeth
Yu *et al.* [65]	MRM	Local methylation profile from both regions and subjects	RRBS	~ 300 K CpGs in the promoter regions of chromosome 17	https://github.com/yuft2003/MRM

method for single-cell DNA methylation profiles [66]. Based on the local DNA sequence and observed neighboring methylation states, DeepCpG constructed a deep neural network model to predict the binary CpG methylation states. It provided insights into how sequence composition affects methylation variability, and enabled accurate imputation of missing methylation states, which provided great facilitation for genome-wide downstream analyses.

Then Jiang *et al*. proposed a novel LightCpG model to identify the DNA methylation status of CpG sites in single cells [67]. Using the sequence features, structure features, positional features, and their combinations, they constructed a novel gradient boosting decision tree (GBDT) algorithm including gradient-based one-side sampling and exclusive feature bundling to reduce the training time. A series of validation experiments demonstrated that the LightCpG model could achieve outstanding performance in the imputation of DNA methylation levels with low computational complexity.

Combined with frameworks of methylation clustering, two groups developed different imputation methods separately. Kapourani and Sanguinetti proposed a Bayesian hierarchical method called Melissa which addressed data imputation on unassayed CpG loci [68]. It effectively apply both the information of neighboring CpGs and of other cells with similar methylation patterns in order to predict CpG methylation states, and output the probability distributions that fully quantify the uncertainty of the methylation prediction. It is more feasible for downstream design and analysis compared to the point-estimates provided by traditional prediction approaches. De Souza *et al*. proposed a novel statistical model and framework called Epiclomal to impute the missing methylation values [69]. By using a hierarchical mixture model to borrow statistical strength across cells and neighboring loci, Epiclomal could impute the inherent missing CpG methylation values more correctly on one hand, and also outperformed previous non-probabilistic methods to find the true clusters on the other hand.

Our group also implemented an imputation method based on the CatBoost gradient boosting model to impute the single-cell methylation states [70]. Besides the DNA sequence information, intracellular neighboring methylation states, and the similarity of local methylation patterns between units, CaMelia was designed to construct a separate model for each cell which could avoid excessive smoothing of all cells and learn the unicellular nature of cells. Experimental validation results on real single-cell methylation datasets indicated that CaMelia yielded significant imputation performance improvement over previous methods. In the downstream analysis of cell-type identification, our CaMelia model could help to discover more intercellular differentially methylated loci that were masked in sparse raw data, and improve the identification of cell subpopulations.

THE IDENTIFICATION OF DIFFERENTIAL METHYLATION PATTERN

The identification of differential methylation pattern (DMP) is an effective way to discover the abnormal methylated genes closely related to tissue specificity and the development of diseases, such as cancer. A comprehensive series of methods identifying the DMP were developed according to the whole genome sequencing data and array data [71–82].

DMP detection for sequencing data

For the sequencing methylation data, the Zhang group developed the first user-friendly tool named CpG_MPs for identification and analysis of the methylation patterns of genomic regions from bisulfite sequencing data [71]. CpG_MPs defined the interested regions using the methylation status of neighboring CpGs by hotspot extension algorithm, and the differentially methylated regions across paired or multiple samples are identified with a combinatorial Shannon entropy algorithm. Applying the tool on human bisulfite sequencing data during cellular differentiation, some potentially functional regions related to cellular differentiation were retrieved.

Then Park *et al.* presented a statistical package named methylSig to analyze the difference between disease and control samples [72]. Taking the read coverage and biological variation into consideration, methlSig applied a beta-binomial approach to identify methylation differences for CpG loci or regions across defined groups. It had high sensitivity and potential to identify the important regions in diseased samples.

The Wang group proposed a novel integrative statistical framework named M&M which combined MeDIP-seq and MRE-seq data to detect differentially methylated regions [73]. By modeling the relationship between DNA methylation level, CpG content, and expected sequencing reads in defined region, M&M constructed a statistical model to compute a probability score and identified the differential methylated regions.

As the methods above did not provide accurate statistical inference, Korthauer *et al.* developed an inferential approach based on a pooled null distribution [74]. Based on a nested autoregressive correlated error structure, it fit a generalized least squares regression model and constructed a region-level statistic to detect

the DMP. The detection results based on both Monte Carlo simulation and experimental data indicated that it could improve the specificity and sensitivity of lists of regions and control the false discovery rate accurately.

DMP detection for array data

For array methylation data, the mean and the median methylation values between the compared groups were applied for the differential methylation detection. The Zhang group developed FastDMA which aimed to identify significantly differentially methylated probes and differentially methylated region [75]. Based on a uniformed statistical model, analysis of covariance was used to analyze the Beadchip array data for the DMP scanning. Many differentially methylated sites in different types of cancer were identified on three large-scale DNA methylation datasets from TCGA.

To better use the correlation of nearby CpG sites, Shen *et al*. presented a novel Bayesian framework named DMRMark for detecting DMRs from methylation array data [76]. By combining the constrained Gaussian mixture model and the biological knowledge with the nonhomogeneous hidden Markov model, DMRMark could detect the DMRs without the requirement of predefined boundaries or decision windows, and could detect DMRs from even a single pair of samples or unpaired samples.

Methods applicable to both of the sequencing and array methylation data were also proposed. Zhang *et al*. developed an unsupervised approach named QDMRs to identify DMRs with Shannon entropy [77]. It is applied to MeDIP-chip data of human tissues and RRBS data of mouse, and identified some potential functional regions, which indicated that QDMR was a platform-free and species-free model. Besides the regions of CpG islands and promoter regions, Lee *et al*. presented a wavelet-based functional mixed model to detect the DMRs in other genomic regions [78]. By accommodating spatial correlations across genomes and correlations between samples through functional random effects, they developed the framework which could be applied to different settings and had more power in the DMR detection. Then Denault *et al*. improved a powerful wavelet-based method called Fast Functional Wavelet detect the DMRs fast [79]. Combining the results of theoretical null distribution of Bayes factors, it achieved fast emulation of the test statistic and reduced the computational time considerably.

DISCUSSION AND FUTURE PERSPECTIVES

More and more studies and results indicated the great importance of DNA methylation changes across many types of solid tumors, hematological malignancies, autoimmune diseases, metabolic and neurological disorders as well as the aging [83–94]. Many aberrant DNA methylation patterns have been demonstrated to be specific of certain diseases, and the presence of differential methylation patterns are potentially diagnostic, prognostic and predictive markers. Although a lot of enormous progress in understanding the role of DNA methylation in different diseases, only few methylated genes or functional elements serve as clinically relevant cancer biomarkers [95]. Therefore, it is urgent to identify the most effective methylation biomarkers for the specific disease diagnosis, prognosis or response to treatment [11,96]. Rapid development of global DNA methylation profiling techniques has allowed our understanding of diverse aspects of genomic DNA methylation in health and disease, the bottleneck in DNA methylation advances has shifted from data generation to data analysis [32].

As the most promising direction for DNA methylation, the current single-cell methylation measurement techniques are seriously plagued by the low coverage of CpGs assayed. Therefore, it is essential to develop versatile imputation tools and address the inherent sparsity of the single-cell methylation data for quantitative analysis of the whole genome. The current imputation models achieved satisfying overall imputation accuracies, however, the biologists are more interested in whether the methylation status of a specific CpG locus could be accurately imputed. Therefore, it would be meaningful to develop the precise imputation model and provide the confidence level for a specific CpG locus of a specific cell.

The detection of differential methylation pattern would identify many abnormal genes/elements in different diseases, but a series of validation experiments are necessary to confirm the accuracy and reproducibility. In the validation, the receiver operator characteristic (ROC) curve was used to determine the sensitivity and specificity of the candidate biomarker genes [97,98], Cox proportional hazard model or Kaplan–Meier survival analysis were applied to discover their recurrence and response to treatment [99–101]. Furthermore, the proper integration of multi-omics is also an effective way for deeper insights into disease etiology [102,103], and machine learning methods offer novel techniques to integrate and analyze the various omics data enabling the discovery of new biomarkers [104,105]. As there are three distinct features for the multi-omics data: the large number of features relative to the number of samples, the strong dependencies between features in different omics, and the intrinsic lower dimensional representation of the omics data,

more versatile and powerful machine learning models need to be carefully developed in the future [106].

ACKNOWLEDGEMENTS

We thank Prof. Wei Wang at UCSD for insightful comments during manuscript preparation. This work was supported by the National Natural Science Foundation of China (No. 61872063) and Shenzhen Science and Technology Program (No. JCYJ20210324140407021).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Shicai Fan, Likun Wang, Liang Liang, Xiaohong Cao, Jianxiong Tang and Qi Tian declare that they have no conflict of interests.

REFERENCES

1. Bird, A. P. (1986) CpG-rich islands and the function of DNA methylation. Nature, 321, 209−213
2. Reik, W. and Walter, J. (2001) Genomic imprinting: parental influence on the genome. Nat. Rev. Genet., 2, 21−32
3. Reik, W., Dean, W. and Walter, J. (2001) Epigenetic reprogramming in mammalian development. Science, 293, 1089−1093
4. Mohandas, T., Sparkes, R. S. and Shapiro, L. J. (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science, 211, 393−396
5. Gartler, S. M. and Riggs, A. D. (1983) Mammalian X-chromosome inactivation. Annu. Rev. Genet., 17, 155−190
6. Reik, W., Collick, A., Norris, M. L., Barton, S. C. and Surani, M. A. (1987) Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature, 328, 248−251
7. Chen, S., Yan, G., Zhang, W., Li, J., Jiang, R. and Lin, Z. (2021) RA3 is a reference-guided approach for epigenetic characterization of single cells. Nat. Commun., 12, 2177
8. Liu, Q., Xia, F., Yin, Q. and Jiang, R. (2018) Chromatin accessibility prediction via a hybrid deep convolutional neural network. Bioinformatics, 34, 732−738
9. Baylin, S. B. and Jones, P. A. (2011) A decade of exploring the cancer epigenome—biological and translational implications. Nat. Rev. Cancer, 11, 726−734
10. Skvortsova, K., Stirzaker, C. and Taberlay, P. (2019) The DNA methylation landscape in cancer. Essays Biochem., 63, 797−811
11. Fan, S. and Chi, W. (2016) Methods for genome-wide DNA methylation analysis in human cancer. Brief. Funct. Genomics, 15, 432−442
12. Laird, P. W. (2010) Principles and challenges of genomewide DNA methylation analysis. Nat. Rev. Genet., 11, 191−203
13. Bock, C. (2012) Analysing and interpreting DNA methylation data. Nat. Rev. Genet., 13, 705−719
14. Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M. and Jacobsen, S. E. (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature, 452, 215−219
15. Friso, S., Choi, S. W., Dolnikowski, G. G. and Selhub, J. (2002) A method to assess genomic DNA methylation using high-performance liquid chromatography/electrospray ionization mass spectrometry. Anal. Chem., 74, 4526−4531
16. Lisanti, S., Omar, W. A., Tomaszewski, B., De Prins, S., Jacobs, G., Koppen, G., Mathers, J. C. and Langie, S. A. (2013) Comparison of methods for quantification of global DNA methylation in human cells and tissues. PLoS One, 8, e79044
17. Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H. and Ecker, J. R. (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell, 133, 523−536
18. Kelly, T. K., Liu, Y., Lay, F. D., Liang, G., Berman, B. P. and Jones, P. A. (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res., 22, 2497−2506
19. Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L. and Schübeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet., 37, 853−862
20. Pidsley, R., Zotenko, E., Peters, T. J., Lawrence, M. G., Risbridger, G. P., Molloy, P., Van Djik, S., Muhlhausler, B., Stirzaker, C. and Clark, S. J. (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol., 17, 208
21. Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R., Down, T. A., *et al*. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet., 38, 1378−1385
22. Tian, Y., Morris, T. J., Webster, A. P., Yang, Z., Beck, S., Feber, A. and Teschendorff, A. E. (2017) ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics, 33, 3982−3984
23. Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S. R., Stegle, O., Reik, W. and Kelsey, G. (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods, 11, 817−820
24. Angermueller, C., Clark, S. J., Lee, H. J., Macaulay, I. C., Teng, M. J., Hu, T. X., Krueger, F., Smallwood, S., Ponting, C. P.,

Voet, T., *et al*. (2016) Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat. Methods, 13, 229−232
25. Pott, S. (2017) Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. eLife, e23203
26. Wang, Y., Wang, A., Liu, Z., Thurman, A. L., Powers, L. S., Zou, M., Zhao, Y., Hefel, A., Li, Y., Zabner, J., *et al*. (2019) Single-molecule long-read sequencing reveals the chromatin basis of gene expression. Genome Res., 29, 1329−1342
27. Gu, H., Raman, A. T., Wang, X., Gaiti, F., Chaligne, R., Mohammad, A. W., Arczewska, A., Smith, Z. D., Landau, D. A., Aryee, M. J., *et al*. (2021) Smart-RRBS for single-cell methylome and transcriptome analysis. Nat. Protoc., 16, 4004−4030
28. Yong, W. S., Hsu, F. M., and Chen, P. Y. (2016) Profiling genome-wide DNA methylation. Epigenet Chromatin. 9, 26
29. Dirks, R. A., Stunnenberg, H. G. and Marks, H. (2016) Genome-wide epigenomic profiling for biomarker discovery. Clin. Epigenetics, 8, 122
30. Rauluseviciute, I., Drabløs, F. and Rye, M. B. (2019) DNA methylation data by sequencing: experimental approaches and recommendations for tools and pipelines for data analysis. Clin. Epigenetics, 11, 193
31. Zuo, T., Tycko, B., Liu, T. M., Lin, J. J. and Huang, T. H. (2009) Methods in DNA methylation profiling. Epigenomics, 1, 331−345
32. Li, S. and Tollefsbol, T. O. (2021) DNA methylation methods: Global DNA methylation and methylomic analyses. Methods, 187, 28−43
33. Arora, I. and Tollefsbol, T. O. (2021) Computational methods and next-generation sequencing approaches to analyze epigenetics data: Profiling of methods and applications. Methods, 187, 92−103
34. Li, D., Zhang, B., Xing, X. and Wang, T. (2015) Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. Methods, 72, 29−40
35. Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., *et al*. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature, 466, 253−257
36. Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., *et al*. (2010) The NIH roadmap epigenomics mapping consortium. Nat. Biotechnol., 28, 1045−1048
37. Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., *et al*. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature, 489, 57−74
38. Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., *et al*. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature, 454, 766−770
39. Roberts, R. J., Carneiro, M. O. and Schatz, M. C. (2013) The advantages of SMRT sequencing. Genome Biol., 14, 405
40. Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L., Schroth, G. P., Gunderson, K. L., *et al*. (2011) High density DNA methylation array with single CpG site resolution. Genomics, 98, 288−295
41. Moran, S., Arribas, C. and Esteller, M. (2016) Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics, 8, 389−399
42. Fan, S., Tang, J., Li, N., Zhao, Y., Ai, R., Zhang, K., Wang, M., Du, W. and Wang, W. (2019) Integrative analysis with expanded DNA methylation data reveals common key regulators and pathways in cancers. NPJ Genom. Med., 4, 2
43. Guo, H., Zhu, P., Wu, X., Li, X., Wen, L. and Tang, F. (2013) Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res., 23, 2126−2135
44. Farlik, M., Sheffield, N. C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J. and Bock, C. (2015) Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Rep., 10, 1386−1397
45. Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., *et al*. (2018) Single-cell DNA methylome sequencing of human preimplantation embryos. Nat. Genet., 50, 12−19
46. Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., *et al*. (2014) The DNA methylation landscape of human early embryos. Nature, 511, 606−610
47. Han, L., Wu, H. J., Zhu, H., Kim, K. Y., Marjani, S. L., Riester, M., Euskirchen, G., Zi, X., Yang, J., Han, J., *et al*. (2017) Bisulfite-independent analysis of CpG island methylation enables genome-scale stratification of single cells. Nucleic Acids Res., 45, e77
48. Chen, P. Y., Cokus, S. J. and Pellegrini, M. (2010) BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics, 11, 203
49. Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M. Q., Chen, P. Y. and Pellegrini, M. (2013) BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics, 14, 774
50. Huang, K. Y. Y., Huang, Y. J. and Chen, P. Y. (2018) BS-Seeker3: ultrafast pipeline for bisulfite sequencing. BMC Bioinformatics, 19, 111
51. Byun, H. M., Siegmund, K. D., Pan, F., Weisenberger, D. J., Kanel, G., Laird, P. W. and Yang, A. S. (2009) Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. Hum. Mol. Genet., 18, 4808−4817
52. Caliskan, M., Cusanovich, D. A., Ober, C. and Gilad, Y. (2012) The effects of EBV transformation on gene expression levels and methylation profiles. Hum. Mol. Genet., 20, 1643−1652
53. Zou, L. S., Erdos, M. R., Taylor, D. L., Chines, P. S., Varshney, A., Parker, S. C. J., Collins, F. S., Didion, J. P. and Inst, M. G., and the McDonnell Genome Institute. (2018) BoostMe

accurately predicts DNA methylation values in whole-genome bisulfite sequencing of multiple human tissues. BMC Genomics, 19, 390

54. Zhang, W., Spector, T. D., Deloukas, P., Bell, J. T. and Engelhardt, B. E. (2015) Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. Genome Biol., 16, 14
55. Fang, F., Fan, S., Zhang, X. and Zhang, M. Q. (2006) Predicting methylation status of CpG islands in the human brain. Bioinformatics, 22, 2204−2209
56. Ma, B., Wilker, E. H., Willis-Owen, S. A. G., Byun, H. M., Wong, K. C. C., Motta, V., Baccarelli, A. A., Schwartz, J., Cookson, W. O. C. M., Khabbaz, K., *et al*. (2014) Predicting DNA methylation level across human tissues. Nucleic Acids Res., 42, 3515−3528
57. Pavlovic, M., Ray, P., Pavlovic, K., Kotamarti, A., Chen, M. and Zhang, M. Q. (2017) DIRECTION: a machine learning framework for predicting and characterizing DNA methylation and hydroxymethylation in mammalian genomes. Bioinformatics, 33, 2986−2994
58. Ma, B., Allard, C., Bouchard, L., Perron, P., Mittleman, M. A., Hivert, M. F. and Liang, L. (2019) Locus-specific DNA methylation prediction in cord blood and placenta. Epigenetics, 14, 405−420
59. Tian, Q., Zou, J., Tang, J., Fang, Y., Yu, Z. and Fan, S. (2019) MRCNN: a deep learning model for regression of genome-wide DNA methylation. BMC Genomics, 20, 192
60. Fan, S., Huang, K., Ai, R., Wang, M. and Wang, W. (2016) Predicting CpG methylation levels by integrating Infinium HumanMethylation450 BeadChip array data. Genomics, 107, 132−137
61. Fan, S., Li, C., Ai, R., Wang, M., Firestein, G. S. and Wang, W. (2016) Computationally expanding infinium HumanMethylation450 BeadChip array data to reveal distinct DNA methylation patterns of rheumatoid arthritis. Bioinformatics, 32, 1773−1778
62. Zheng, Y., Joyce, B. T., Liu, L., Zhang, Z., Kibbe, W. A., Zhang, W. and Hou, L. (2017) Prediction of genome-wide DNA methylation in repetitive elements. Nucleic Acids Res., 45, 8697−8711
63. Li, G., Raffield, L., Logue, M., Miller, M. W., Santos, H. P., O' Shea, T. M., Fry, R. C. and Li, Y. (2020) CUE: CpG imputation ensemble for DNA methylation levels across the human methylation450 (hm450) and epic (hm850) beadchip platforms. Epigenetics, 16, 851−861
64. Tang, J., Zou, J., Zhang, X., Fan, M., Tian, Q., Fu, S., Gao, S. and Fan, S. (2020) PretiMeth: precise prediction models for DNA methylation based on single methylation mark. BMC Genomics, 21, 364
65. Yu, F., Xu, C., Deng, H. W. and Shen, H. (2020) A novel computational strategy for DNA methylation imputation using mixture regression model (MRM). BMC Bioinformatics, 21, 552
66. Angermueller, C., Lee, H. J., Reik, W. and Stegle, O. (2017) DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. Genome Biol., 18, 67
67. Jiang, L., Wang, C., Tang, J. and Guo, F. (2019) LightCpG: a multi-view CpG sites detection on single-cell whole genome sequence data. BMC Genomics, 20, 306
68. Kapourani, C. A. and Sanguinetti, G. (2019) Melissa: Bayesian clustering and imputation of single-cell methylomes. Genome Biol., 20, 61
69. P E de Souza, C., Andronescu, M., Masud, T., Kabeer, F., Biele, J., Laks, E., Lai, D., Ye, P., Brimhall, J., Wang, B., *et al*. (2020) Epiclomal: Probabilistic clustering of sparse single-cell DNA methylation data. PLOS Comput. Biol., 16, e1008270
70. Tang, J., Zou, J., Fan, M., Tian, Q., Zhang, J. and Fan, S. (2021) CaMelia: imputation in single-cell methylomes based on local similarities between cells. Bioinformatics, 37, btab029
71. Su, J., Yan, H., Wei, Y., Liu, H., Liu, H., Wang, F., Lv, J., Wu, Q. and Zhang, Y. (2013) CpG_MPs: identification of CpG methylation patterns of genomic regions from high-throughput bisulfite sequencing data. Nucleic Acids Res., 41, e4
72. Park, Y., Figueroa, M. E., Rozek, L. S. and Sartor, M. A. (2014) MethylSig: a whole genome DNA methylation analysis pipeline. Bioinformatics, 30, 2414−2422
73. Zhang, B., Zhou, Y., Lin, N., Lowdon, R. F., Hong, C., Nagarajan, R. P., Cheng, J. B., Li, D., Stevens, M., Lee, H. J., *et al*. (2013) Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. Genome Res., 23, 1522−1540
74. Korthauer, K., Chakraborty, S., Benjamini, Y. and Irizarry, R. A. (2019) Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. Biostatistics, 20, 367−383
75. Wu, D., Gu, J. and Zhang, M. Q. (2013) FastDMA: an infinium humanmethylation450 beadchip analyzer. PLoS One, 8, e74275
76. Shen, L., Zhu, J., Robert Li, S. Y. and Fan, X. (2017) Detect differentially methylated regions using non-homogeneous hidden Markov model for methylation array data. Bioinformatics, 33, 3701−3708
77. Zhang, Y., Liu, H., Lv, J., Xiao, X., Zhu, J., Liu, X., Su, J., Li, X., Wu, Q., Wang, F., *et al*. (2011) QDMR: a quantitative method for identification of differentially methylated regions by entropy. Nucleic Acids Res., 39, e58
78. Lee, W. and Morris, J. S. (2016) Identification of differentially methylated loci using wavelet-based functional mixed models. Bioinformatics, 32, 664−672
79. Denault, W. R. P. and Jugessur, A. (2021) Detecting differentially methylated regions using a fast wavelet-based approach to functional association analysis. BMC Bioinformatics, BMC Bioinformatics. 22, 61. doi: 10.1186/s12859–021-03979-y
80. Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., Qin, Z., Jin, P. and Conneely, K. N. (2015) Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. Nucleic Acids Res., 43, e141
81. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A. and Mason, C. E. (2012) methylKit: a comprehensive R package for the analysis of

genome-wide DNA methylation profiles. Genome Biol., 13, R87
82. Feng, H. and Wu, H. (2019) Differential methylation analysis for bisulfite sequencing using DSS. Quant. Biol., 7, 327−334
83. Nishiyama, A. and Nakanishi, M. (2021) Navigating the DNA methylation landscape of cancer. Trends Genet., 37, 1012−1027
84. Feinberg, A. P., Koldobskiy, M. A. and Göndör, A. (2016) Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. Nat. Rev. Genet., 17, 284−299
85. Hanahan, D. and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. Cell, 144, 646−674
86. McDonald, O. G., Li, X., Saunders, T., Tryggvadottir, R., Mentch, S. J., Warmoes, M. O., Word, A. E., Carrer, A., Salz, T. H., Natsume, S., *et al*. (2017) Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. Nat. Genet., 49, 367−376
87. Glossop, J. R., Nixon, N. B., Emes, R. D., Sim, J., Packham, J. C., Mattey, D. L., Farrell, W. E. and Fryer, A. A. (2017) DNA methylation at diagnosis is associated with response to disease-modifying drugs in early rheumatoid arthritis. Epigenomics, 9, 419−428
88. Sun, Z. H., Liu, Y. H., Liu, J. D., Xu, D. D., Li, X. F., Meng, X. M., Ma, T. T., Huang, C. and Li, J. (2017) Mecp2 regulates ptch1 expression through DNA methylation in rheumatoid arthritis. Inflammation, 40, 1497−1508
89. Ai, R., Hammaker, D., Boyle, D. L., Morgan, R., Walsh, A. M., Fan, S., Firestein, G. S. and Wang, W. (2016) Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes. Nat. Commun., 7, 11849
90. Ai, R., Laragione, T., Hammaker, D., Boyle, D. L., Wildberg, A., Maeshima, K., Palescandolo, E., Krishna, V., Pocalyko, D., Whitaker, J. W., *et al*. (2018) Comprehensive epigenetic landscape of rheumatoid arthritis fibroblast-like synoviocytes. Nat. Commun., 9, 1921
91. Braun, K. V. E., Dhana, K., de Vries, P. S., Voortman, T., van Meurs, J. B. J., Uitterlinden, A. G., Hofman, A., Hu, F. B., Franco, O. H. and Dehghan, A. (2017) Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin. Epigenetics, 9, 15
92. Liu, Z., Li, X., Zhang, J. T., Cai, Y. J., Cheng, T. L., Cheng, C., Wang, Y., Zhang, C. C., Nie, Y. H., Chen, Z. F., *et al*. (2016) Autism-like behaviours and germline transmission in transgenic monkeys overexpressing MeCP2. Nature, 530, 98−102
93. Eryilmaz, I. E., Cecener, G., Erer, S., Egeli, U., Tunca, B., Zarifoglu, M., Elibol, B., Bora Tokcaer, A., Saka, E., Demirkiran, M., *et al*. (2017) Epigenetic approach to early-onset Parkinson's disease: low methylation status of SNCA and PARK2 promoter regions. Neurol. Res., 39, 965−972
94. Horvath, S. (2013) DNA methylation age of human tissues and cell types. Genome Biol., 14, R115
95. Pajares, M. J., Palanca-Ballester, C., Urtasun, R., Alemany-Cosme, E., Lahoz, A. and Sandoval, J. (2021) Methods for analysis of specific DNA methylation status. Methods, 187, 3−12
96. Mallik, S. and Zhao, Z. (2017) Towards integrated oncogenic marker recognition through mutual information-based statistically significant feature extraction: an association rule mining based study on cancer expression and methylation profiles. Quant. Biol., 5, 302−327
97. van den Helder, R., Wever, B. M. M., van Trommel, N. E., van Splunter, A. P., Mom, C. H., Kasius, J. C., Bleeker, M. C. G. and Steenbergen, R. D. M. (2020) Non-invasive detection of endometrial cancer by DNA methylation analysis in urine. Clin. Epigenetics, 12, 165
98. Wentzensen, N., Bakkum-Gamez, J. N., Killian, J. K., Sampson, J., Guido, R., Glass, A., Adams, L., Luhn, P., Brinton, L. A., Rush, B., *et al*. (2014) Discovery and validation of methylation markers for endometrial cancer. Int. J. Cancer, 135, 1860−1868
99. Mao, Y. K., Liu, Z. B. and Cai, L. (2020) Identification of glioblastoma-specific prognostic biomarkers via an integrative analysis of DNA methylation and gene expression. Oncol. Lett., 20, 1619−1628
100. Zhao, J., Wang, L., Kong, D., Hu, G. and Wei, B. (2020) Construction of novel DNA methylation-based prognostic model to predict survival in glioblastoma. J. Comput. Biol., 27, 718−728
101. Harada, H., Miyamoto, K., Yamashita, Y., Nakano, K., Taniyama, K., Miyata, Y., Ohdan, H. and Okada, M. (2013) Methylation of breast cancer susceptibility gene 1 (BRCA1) predicts recurrence in patients with curatively resected stage I non-small cell lung cancer. Cancer, 119, 792−798
102. Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S. 2nd and Byrum, S. D. (2021) Multi-omics data integration considerations and study design for biological systems and disease. Mol. Omics, 17, 170−185
103. Cao, S., Zhao, Y., Wu, Y., Song, T., Burair, A. and Xu, Y. (2017) Transcription regulation by DNA methylation under stressful conditions in human cancer. Quant. Biol., 5, 328−337
104. Reel, P. S., Reel, S., Pearson, E., Trucco, E. and Jefferson, E. (2021) Using machine learning approaches for multi-omics data analysis: A review. Biotechnol. Adv., 49, 107739
105. Ma, T. and Zhang, A. (2019) Integrate multi-omics data with biological interaction networks using Multi-view Factorization AutoEncoder (MAE). BMC Genomics, 20, 944
106. Rappoport, N. and Shamir, R. (2018) Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Res., 46, 10546−10562