

## DNA Methylation Analysis Methods for Cancer Research

Kai Zheng<sup>1</sup>, Hanzhe Liu<sup>2</sup>, Simin Zhu<sup>1</sup>, Huamei Li<sup>1</sup> and Xiaozhou Chen<sup>1,\*</sup>

<sup>1</sup>Key Laboratory of IOT Application Technology of Universities in Yunnan Province,  
Yunnan Minzu University, Kunming 650031, China

<sup>2</sup>School of Electrical and Information Technology, Yunnan Minzu University, Kunming 650031, China

\*Corresponding author, E-mail: ch\_xiaozhou@163.com

**Keywords:** DNA Methylation; DNA Methyltransferases; Methylation Detection; Cancer

**Abstract.** DNA methylation is an important epigenetic modification mechanism, interacting with histone modification, changing the structure of chromatin and regulating gene expression. Meanwhile, there is a close relationship with cell proliferation, aging, cancer and other life phenomena. DNA methylation includes DNA undermethylation and local high CpG island methylation on the whole genome. By the help of DNA methyltransferases (DNMTs), DNA methylation can be completed. With the development of further research on epigenetics and DNA methylation, genomic methylation detection technology has got rapid development. DNA methylation detection methods can be divided into the following categories: methylation detection, gene-specific methylation detection and searching new methylation sites. Researches on DNA methylation showed that abnormal genomic DNA methylation and tumor is closely related to the occurrence and development of DNA hypomethylation and undermethylation, which can cause abnormal expression of genes and lead to cancer. At the same time, the change of DNA methylation also provides a new way for tumor detection and treatment, which provides a new way for the treatment of cancer. In this paper, the current main analysis methods and research progress of DNA methylation were summarized and discussed preliminarily. Moreover, the relationship of abnormal DNA methylation with the occurrence and development of cancer were reviewed.

### Introduction

Under the influence of DNA methylation transferases, the 5 'C is translated into 5'mC in the CpG island ( DNA fragments about 300 ~ 3000 bp that contain CpG dideoxynucleotide, which is often located in the first exon region and the promoter region with the length of 1 kb) [1], and the way of DNA modification does not change the gene sequences, but it can regulate and control the gene expression, especially interact with histone modification, change the chromatin structure and regulate the gene expression [2]. By regulating the expression of genes, these epigenetics modifications form unique phenotypes in the cells and affect the cell genes expression, which affection will be passed to the next generation with cell division. In mammals, DNA methylation mostly appears in the 5 'C of cytosine, the vast majority of which are located in the CpG island, about 2000 to 4000 CpG islands being CpG sites and about 70% being mCpG sites [3]. The CpG sites are normal in healthy people, but those out of the CpG island are methylated usually. Methylation can keep stable in the process of cell proliferation. In cancer cells, the methylation level of CpG sequences that out of the tumor suppressor gene CpG island are reduced, but the gene sequences of CpG in the CpG island are highly methylated, which increases chromosome spiral, and the expression of tumor suppressor genes are restrained [4].

Eukaryotic genomic DNA methylation mainly catalyzed by three kinds of DNA methyltransferases: DNMT1, DNMT2 and DNMT3 (including DNMT3a and DNMT3b). DNMT1 is the main type of mammal DNMTs, which main function is maintaining methylation, existing in nearly all of the cells and abundant in the proliferation cells. Although DNMT2 has basic sequences of methyl transfer activity, it was not found in the cell currently. DNMT3 is responsible for de novo

methylation, which presents high expression in the embryonic stem cells and the early embryonic cells (de novo methylation is the methylation process in the unmethylated double-stranded DNA). DNMT1 is isolated from mice firstly, expressing in the reproductive cells widely, locating in DNA replication forks, and it is associated with proliferating cellular nuclear antigen (PCNA), histone deacetylase 2 (HDAC2), and DNAP1 (protein related to DNMT1), then becoming complex. DNMT1 maintains the sequence CG methylation. DNMT3 was found in the body of mice, people and zebrafish including DNMT3a and DNMT3b which are abundant in undifferentiated embryonic stem cells but the amount is small in body cells. Although the main function is de novo methylation, to a certain extent, it also plays an important role in maintaining DNA methylation.

Methyl group is transferred to cytosine by DNA methyltransferase (DNMT) catalyzing, which is involved in forming and maintaining methylation, however, the effect of demethylation enzyme is opposite. Usually, the CpG island is the requirement of gene transcription, existing in gene promoter region, if it is methylated, gene transcription will be reined in. Research showed that the distribution and the level of related epigenetic markers have significant change in the cancer genome, which suggests that the change of epigenetic markers may lead to the formation of tumor [5]. Promoter region methylation can prevent gene expression and DNA methylation abnormalities may be the cause of some certain cancers, for example, high or low methylation in gene promoter region can lead to some tumor suppressor genes silence or some genes expression abnormal, resulting in the cell or the organization canceration.

## **Methods of DNA methylation Detection in Genome**

### **The method of restriction enzyme that is sensitive to methylation.**

Methylation sensitivity restriction enzymes will cut the unmethylated sites, but methylated sites can't be cut, DNA sequences will be cut into different sizes of fragments, and then they can be analyzed. There are several methylation sensitivity restriction enzymes commonly used: Hpa II -Msp I and Sma I -XmaI, the former can identify sequence CCGG and the later can identify sequence CCCGGG. For the number of Sma I -XmaI is more than Hpa II -Msp I and low frequency in the body, Hpa II -Msp I is used widely. Hpa II and Msp I can identify sequence CCGG, but Hpa II does not cut methylated cytosine of DNA sequences, so Hpa II - Msp I is used to process DNA sequences, and the PCR amplification results are separated to determine the methylation and unmethylation status [6]. This method can detect several CpG sites in the genome at the same time, and the results of the experiment are easy to analysis and low cost. As a result of CG is not limited to sequence CCGG, CG that is not in this sequence will be ignored. In addition, if the sites are methylated in transcription process, the test results will be meaningful, and the test method needs a large sample so that it is not suitable for mixed samples. Moreover, the incomplete enzyme cutting can cause false positive results.

### **The method of PCR based on DNA processed by Bisulfite.**

The principle of the method is that the unmethylated cytosine can be translated into uracil after processed by Bisulfite, then uracil will be translated into thymine by PCR amplification, but methylated cytosine is unchanged. First of all, the genomic DNA should be processed by MspI, and then recycling a certain length of segments to be processed by Bisulfite. The DNA duplexes are no longer complementary, combining this method with high-throughput sequencing technologies can detect methylation status of a single base site. In mammals methylation databases, methylated cytosine sites are suited to the binomial distribution by single site methylation data of BS. Binomial distribution test according to the error rate to determine whether a site is methylated or not. The quality value of the base sequencing that compared to the sites of the short sequences being higher, the cytosine being more and the error rate being lower, so the probability of methylation site is higher. Therefore, the cytosine site with high methylation rate is easy to be identified as methylation site. The binomial distribution test formula is as follows:

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (1)$$

where

$n$  = the sequencing depth of a certain site;

$x$  = the number of bases compared to the site but cytosine;

$P$  = the error rate.

The transformation of the calculation formula is as follows:

$$T_p = (1-r) \frac{N_{CpG}}{N_{mCpG}} \times 100\% \quad (2)$$

where

$T_p$  = the ratio of the number of nonCG that quality value is greater than 20 and the total number of nonCG;

$N_{CpG}$  = the total number of CpG islands;

$N_{mCpG}$  = the number of methylation C of CpG islands.

Although the methylated cytosine also exists in nonCG with low frequency, it still can raise the false positive rate. In addition, amplifying fragments to be tested by the designed BSP primer, all uracils become thymines, then detecting the sequences of PCR results can determine whether the sites are methylated or not. Therefore, Bisulfite processing is the "gold standard" to test the sequences of DNA cytosine methylation. According to this principle, not only can we assess the cytosine methylation sites in the DNA sequence accurately, but we can make quantitative analysis on a single site methylation. With high detection accuracy, it can detect the methylation status of CpG sites of the fragments to be tested, but a lot of cloning and sequencing operation are necessary, tedious and costly. Additionally, for the process involving fragments recycle, the coverage is limited on the genome of this method. With Bisulfite processing, DNA sequence complexity is reduced observably, which is adverse to data processing subsequently.

#### **Methylated DNA immunoprecipitation sequencing.**

Methylated DNA immunoprecipitation sequencing (MeDIP-Seq) is a method of high-throughput sequencing by using of the high methylated DNA fragments of 5'mC antibody abundant in genome [7]. The amount of data used in this method is small, but the data cover widely. MeDIP-Seq can also be used to detect methylated sites in genome-wide with higher cost performance than processed by Bisulfite and it can be used to compare and analyze methylated DNA sites in many samples. For this method is more sensitive to high CpG density region [8], antibodies are enriched in high methylated CpG region with preference and this method can search genomics for methylation sites quickly, accurately and efficiently. Compared DNA methylation patterns in different organizations, which can be used in the study of molecular biology and the related diseases.

#### **High performance liquid chromatography (HPLC) testing technology.**

In 1980s, Kuo and others put forward high performance liquid chromatography (HPLC) method to detect genome-wide DNA methylation level for the first time. DNA samples were hydrolyzed to bases by hydrochloric acid and hydrofluoric acid firstly. Secondly, measuring the ultraviolet absorption peaks of the hydrolysate by chromatographic column and comparing them with standard bases samples. Finally calculating the integral of 5'mC/ (5'mC + 5'C) in the whole genome methylation level.

$$S = \int \frac{5'mC}{5'mC + 5'C} dC \quad (3)$$

Oefner and others proposed denatured high performance liquid chromatography (DHPLC) can be used in the analysis of single nucleotide and DNA molecules. Deng Dajun improved and combined it

with PCR and established a DHPLC detection mechanism to analyze the level of methylation. Firstly, the results of the Bisulfite processing needed to be amplified dissimilarly, mC was not changed after processed by Bisulfite, therefore, higher temperature was needed for its denaturation in the PCR amplification, then PCR results stay for a long time in the chromatographic column, so the methylation level can be detected.

#### **Luminometric methylation assay.**

Karimiet and others found genome-wide DNA luminometric methylation assay (LUMA) firstly. The detection technology referring to EcoR I, processing DNA in two independent reactions (Msp I +EcoR and Hpa II +EcoR I), in which isoschizomer Msp I /Hpa II is commonly used in DNA methylation detection. The 5'end become "- AATT -" after digested by restriction enzyme EcoR I and Msp I /Hpa II become "- GC -". Amplificating polymerase chains and filling the nucleotides to the 5'end [9]. So the methylation level of Msp I /Hpa II sites in the genome can be quantified as the genome-wide methylation percentage:

$$mP = \frac{1 - (\text{Hpa II } \Sigma G / \Sigma T)}{\text{Msp I } \Sigma G / \Sigma T} \times 100\% \quad (4)$$

Using another isoschizomers may also refer to this method for quantitative detection to assess the genome-wide methylation level [8]. Research shows that the methylation probabilities of Sat-alpha, LINE-1 and ALU are high within the scope of the whole genome, accounting for 4%, 11% and 17%. If the number of repeated element is larger than one-third of the CpG sites of the whole genome, it also can be used as alternate index to detect DNA methylation. Such detection method can provide complete 'map' for the whole genome DNA methylation detection, which has been verified in practice.

#### **Pyrosequencing technology.**

Pyrosequencing is a high-throughput DNA sequencing technology based on chemiluminescence quantitative detection technology to measure pyrophosphate, which can detect dozens of samples at the same time. C become U and mC was no changed by Bisulfite processing, therefore, the original methylation sites become the TC single nucleotide polymorphism sites after PCR amplification, the allele frequency of C reflects the genomic DNA methylation level. Followed by cascaded chemiluminescence reaction by four kinds of enzyme catalysis. After annealing the primer and template DNA, then by the influence of DNA polymerase, ATP sulfurylase, apyrase and luciferase, each dNTP polymerization in the primer will be accompanied by one of the release of the fluorescent signal under the action of the DNA polymerase. The fluorescence signal intensity is proportional to the number of dNTP polymerization, then according to the type of dNTP and fluorescence signal intensity, we will get the original genome DNA bases sequences [10]. The detection technology is simple, rapid, high-throughput and quantitative, but only limited length of DNA fragments can be detected at once.

#### **Methylation specific oligonucleotide microarray detection technology.**

Methylation specific oligonucleotide (MSO) microarray is designed for CpG dinucleotide sites methylation specific oligonucleotide. The detection technique requires designing a pair of probes that contain two bases which are not adjacent -- GC (AC). MSO microarray is used to detect whether the sequence is methylated or not. One probe containing GC (5'-GC-GC-3') can identify methylation sequences, and the other probe containing AC (5'-AC-AC-3') can identify the unmethylated sequences, the 5'end of the probe is fixed on the glass by the primer. Firstly, processing DNA fragments with Bisulfite, followed by PCR amplification, the 3'end of amplification results marked with fluorescent, then the amplification results are moved to hybridize on a glass plate with probes. Finally, judge the degree of methylation in sequences according to the fluorescent signal. This method needs to set up a control group. Hence, methylation specific oligonucleotide microarray detection technology can be used for multiple samples, multiple sites methylation detection, which is a high-throughput, high sensitivity and efficient methylation detection method.

## **Methylation and Cancer**

### **The relation of DNA methylation and cancer.**

The CpG islands of gene promoter regions are not methylated normally. Once methylated, the methylation in CpG islands can cause gene silencing and make some important genes such as DNA repair genes, tumor suppressor genes not work. Then DNA damage can not be repaired and cell growth abnormal, which are closely related with the formation of cancer. The undermethylation and local hypermethylation of cancer genome are typical features of cancer.

Local hypermethylation is abnormal methylation in tumor suppressor gene promoter regions usually. Undermethylation in genome-wide is closely related with DNA repetitive sequences and low methylation in promoter regions. Analyze methylation level of several related tumor cell lines of human gene promoter regions (cell lines: colorectal cancer, lung cancer, breast cancer, leukemia, melanoma, glioma, kidney cancer, thyroid cancer and lymphoma). Related genes are P14, P15, P16, O6 MGMT, hMLH1, BRCA1, TIMP-3 and RARB2. Research shows that each specific tumor has one gene promoter region being hypermethylation region at least, and the methylation in the gene promoter region has specificity. For example, the promoter regions of gene O6-MGMT and hMLH1 in colorectal cancer cell lines are hypermethylated, but only promoter region O6-MGMT is hypermethylated in breast cancer cell lines. Bibikova and others clustered the methylation status of lung, breast, colon, prostate cancer cells and healthy cells about 1536 sites in 371 genes. The analysis result indicated that cancer cells and healthy cells as well as various types of cancer cells can be identified in 55 sites [11]. The experiment shows that various cancers have their specific DNA methylation markers.

### **DNA methylation and cancer diagnosis.**

Unlike genetic mutations, epigenetic changes are reversible mostly, researching and developing related drugs to alter epigenetics status may be a new method to cancer treatment. Characteristics of epigenetic changes are associated with the occurrence of cancer cells usually. Therefore, the abnormal cell genome DNA methylation can be used as an important index in early diagnosis of cancer. Methylation specific PCR (MSP) makes the level analysis of cell DNA methylation simple and convenient. In addition, for the sensitivity of MSP is high, it is convenient for the early detection and diagnosis of cancer. The stability of the DNA methylation detection technology and organization specificity are preferable, the degree of change is often associated with the development of cancer. The changes of related genes methylation occurring in many kinds of cancers. For example, peripheral blood cells DNA undermethylation is closely related with the incidence of bladder cancer [12]. The level of methylation of genes CDH13, MYOD1, MGMT, P16INK4b and RASSF1A are changed significantly in bladder cancer, rectal cancer, lung cancer and liver cancer cells.

## **Summary**

The research methods of genomic DNA methylation are varied, but each method has its advantages and disadvantages. Therefore, the appropriate research methods should be selected for different research objects. First of all, a clear research objective is needed, whether the object is the overall level of the genomic DNA methylation or the specific sites of methylation or finding methylation sites in genome-wide. Then select the appropriate method according to actual needs, whether the target sequence is known, whether quantitative research or qualitative research is needed, whether high-throughput detection method and what about the sample source and quantity is needed. Finally, analyze the research object comprehensively and select sensitive, convenient and cost-effective method to process. Along with the progress of technology, there are more and more genome DNA methylation detection methods will be proposed, and the gene chip technology in the detection of high-throughput will be used more widely in the study of genome DNA.

**Acknowledgements:** This research is supported by Yunnan Minzu University Graduate Innovation Fund(Grant No. 2014YJY64)

## References

- [1] Sproul D, Meehan RR. Genomic insights into cancer-associated aberrant CpG island hypermethylation[J]. *Brief Funct Genomics*, 2013, 12(3): 174-90.
- [2] Dahl C, Guldberg P. DNA methylation analysis techniques[J]. *Biogerontology*, 2003, 4: 233-50. Dennis C. Epigenetics and disease: Altered states[J]. *Nature*, 2003, 421(6924): 686-688.
- [3] Jones P A, Baylin S B. The fundamental role of epigenetic events in carcinoma[J]. *NatRev Genet*, 2002, 3(6): 415-428.
- [4] ZHANG Li-Li, WU Jian-Xin. DNA methylation: an epigenetic mechanism for tumorigenesis. *Hereditas(Beijing)*, 2006, 28(7): 880-885.
- [5] Butcher LM, Beck S. AutoMeDIP-seq: a high-throughput, whole genome, DNA methylation assay. *Methods*, 2010, 52(3): 223-231.
- [6] Palmke N, Santacruz D, Walter J. Comprehensive analysis of DNA-methylation in mammalian tissues using MeDIP-chi. *Methods*, 2011, 53(2): 175-184.
- [7] Karimi M, Johansson S, Ekstrom TJ. Using LUMA: a Luminometric-based assay for global DNA-methylation[J]. *Epigenetics*, 2006, 1(1): 45-48.
- [8] Xu X, Gammon MD, Hernandez-Vargas H, et al. DNA methylation in peripheral blood measured by LUMA is associated with breast cancer in a population-based study[J]. *FASEB J*, 2012, 26(6): 2657-2666.
- [9] Bibikova M, Lin Z, Zhou L, et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res*, 2006, 16(3): 383-393.
- [10] Moore L E, Pfeiffer R M, Poscablo C, et al. Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study[J]. *Lancet Oncol*, 2008, 9 (4) : 359-366.
- [11] Yu J, Ni M, Xu J, et al. Methylation profiling of twenty promoter-CpG islands of genes which may contribute to hepatocellular carcinogenesis[J]. *BMC Cancer*, 2002, 2 (1) : 29.