Cancer Letters xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Mini-review

Analyzing the cancer methylome through targeted bisulfite sequencing

Eun-Joon Lee, Junfeng Luo, James M. Wilson, Huidong Shi *

GHSU Cancer Center, Department of Biochemistry and Molecular Biology, Georgia Health Sciences University, Augusta, GA 30912, USA

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
DNA methylation
Targeted bisulfite sequencing
Next-generation sequencing

ABSTRACT

Bisulfite conversion of genomic DNA combined with next-generation sequencing (NGS) has become a very effective approach for mapping the whole-genome and sub-genome wide DNA methylation landscapes. However, whole methylome shotgun bisulfite sequencing is still expensive and not suitable for analyzing large numbers of human cancer specimens. Recent advances in the development of targeted bisulfite sequencing approaches offer several attractive alternatives. The characteristics and applications of these methods are discussed in this review article. In addition, the bioinformatic tools that can be used for sequence capture probe design as well as downstream sequence analyses are also addressed.

1. Introduction

DNA methylation occurs at the carbon-5 position of cytosine residue of the CpG dinucleotide in mammalian genomes and is one of the best-studied epigenetic modifications. DNA methylation is known to play important roles in several key physiological processes including regulation of gene expression, X-chromosome inactivation, imprinting, and maintenance of chromosomal stability. The majority of CpG dinucleotides (70–80%) are fully methylated in normal human cells [1]. However, 0.2–3 kb stretches of GC-rich DNA, called CpG islands (CGIs), appear to be protected from modification in normal somatic cells [2]. Nearly half of all known human genes are associated with CGIs in their 5′-end regulatory regions. Patterns of DNA methylation are stably maintained through somatic cell division and can be inherited across generations. However, studies also show that DNA methylation is dynamically regulated during differentiation and aging [3,4]. A growing body of evidence has suggested that DNA methylation patterns can be modulated by environmental factors [5,6]. Twins whose separation resulted in exposure to variable environmental factors showed significant differences with respect to DNA methylation content [7]. Aberrant DNA methylation changes can cause a number of human diseases such as developmental diseases (ICF syndrome, Prader-Willi and Angelman syndromes, etc.), aging related diseases (i.e. Alzheimer's disease), heart disease, diabetes, and autoimmune diseases [8–12]. Therefore, it is very important to understand how methylation patterns are established and maintained during normal development and under pathological conditions.

Accumulating evidence has indicated that epigenetic alterations are at least as common as, if not more frequent than, mutational events in the development of cancer [13]. Compared to normal cells, the malignant cells exhibit overall genomic hypomethylation (primarily in repeat elements and pericentromeric regions), but simultaneously show hypermethylation of normally protected CGIs [14,15]. Hypermethylation within promoters serves to turn off critical tumor suppressor genes that could otherwise suppress tumorigenesis [15]. Given their important functions in cancer initiation and progression, aberrant methylation patterns may be used as biomarkers for diagnosis and prognosis of cancer [16]. Unlike genetic mutations, epigenetic alterations are reversible, making them a therapeutic target. Treatment with inhibitors of DNA methylation and histone deacetylation can reactivate epigenetically silenced genes and has been shown to restore normal gene function [8,17]. The demethylating agent azacitidine and its deoxy derivative, decitabine, have been approved for treating hematological malignancies including MDS and AML [18].

Abbreviations: ACBS-seq, array capture bisulfite sequencing; AML, acute myeloid leukemia; BSPP, bisulfite padlock probes; CGI, CpG island; COBRA, combined bisulfite and restriction analysis; DMH, differential methylation hybridization; DREAM, digital restriction enzyme analysis of methylation; ES, embryonic stem; HELP, HpaII tiny fragment enrichment by ligation-mediated PCR; ICF, immunodeficiency, centromere instability and facial anomalies; MALDI-TOF MS, matrix-assisted laser desorption/ionization-time of flight mass spectrometry; MBD, methyl-binding domain; MCA, methylated CpG island amplification; MDS, myelodysplastic syndromes; MeDIP-seq, methylated DNA immunoprecipitation sequencing; MeDIP, methylated DNA immunoprecipitation; MIRA-seq, methylated CpG island recovery assay sequencing; MIRA, methylated CpG island recovery assay; MRE-seq, methylation sensitive restriction enzyme sequencing; MSP, methylation specific PCR; NGS, next-generation sequencing; PCR, polymerase chain reaction; SHBS-seq, solution hybrid selection bisulfite sequencing; SHS, solution hybrid selection; TBS-seq, targeted bisulfite sequencing; WGBS-seq, whole genome shotgun bisulfite sequencing.

* Corresponding author. Address: CN-2139, GHSU Cancer Center, 1120 15th Street, Augusta, GA 30912, USA. Tel.: +1 706 721 6000; fax: +1 706 721 2928.
E-mail address: hshi@georgiahealth.edu (H. Shi).

0304-3835/$ - see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.canlet.2012.10.040

2. Methods for DNA methylation analysis

The methylation status of DNA can be analyzed by many different methods which utilize three basic principles: (1) digest unmethylated or methylated DNA with methylation sensitive restriction enzymes; (2) use anti-methylcytosine antibodies or methyl-binding domain (MBD) proteins to enrich methylated DNA; (3) bisulfite treatment can convert unmethylated cytosine into uracil while leaving methylated cytosine unchanged. These principles have been integrated into high-throughput analytical applications such as microarray and next-generation sequencing (NGS) platforms. Several excellent reviews of the genome-wide DNA methylation analysis are available [19–22]. The representative microarray-based methods include DMH [23], HELP [24], MCA [25] (restriction enzyme based), MeDIP [26], and MIRA [27] (affinity enrichment based), which utilize promoter or CpG island tiling arrays to identify differentially methylated genes. Recently, some of these approaches have been moved to the NGS platforms (i.e. HELP-tagging [28], DREAM[29], MRE-seq [30], MeDIP-seq [30], and MIRA-seq [31]). One of the disadvantages of microarray-based analyses, particularly those using enrichment of methylated DNA, is that it can survey for the presence or absence of methylated DNA, but it gives little detail about the extent and pattern of CpG methylation within a given region. Often, extensive follow-up studies must be conducted on a single gene basis to confirm the results of such microarray experiments. The Illlumina Infinium beads array, which analyzes bisulfite-converted DNA, has become a popular platform for methylome analysis. The current version of Infinium 450 K BeadChip is capable of measuring more than 450,000 CpG sites in promoters, CGIs, CGI shores, enhancer regions, and DNase I hypersensitive sites in the human genome curated from published literature. The chip has been proven to be highly accurate, reproducible and does not require a large amount of input DNA [32]. Most importantly, the data analysis is relative simple and straightforward making the Infinium array the best approach currently available for population-based epigenetic studies.

Due to the significantly increased throughput, a large number of differentially methylated genes can now be identified in a single experiment. Therefore, the traditional methods for validation experiments such as MSP, COBRA, and bisulfite clone sequencing are no longer sufficient to keep up with the increased demand. Pyrosequencing (PyroMarkTM, Qiagen) and MassArray (EpiTYPERTM, Sequenom) have now largely replaced the more traditional validation methods. Bisulfite pyrosequencing is a quantitative method for DNA methylation analysis, and it can determine the CpG methylation status in a sequence up to 100-bp in length. Pyrosequencing is based on the sequence-by-synthesis principle; nucleotides are added in each pyrosequencing cycle and the amount of incorporated nucleotide results in a proportional emission of light. DNA methylation ratios are calculated from the levels of light emitted from each nucleotide incorporated at individual CpG positions in a strand-dependent manner. The methylation levels at each CpG position assayed are the average of all amplification products generated during a bisulfite-PCR reaction. DNA methylation analysis by MALDI-TOF mass spectrometry (MS) employs base-specific cleavage of single-stranded nucleic acids [33]. The idea is to generate a PCR product from bisulfite-treated DNA, which is then transcribed *in vitro* into a single-stranded RNA molecule; the synthesized RNA is cleaved in a base-specific manner by an endoribonuclease [33]. The base-specific cleavage products are analyzed using the MassArray instrument. Due to the conversion of unmethylated cytosine to thymine after bisulfite treatment and PCR, different cleavage products are generated for methylated and unmethylated DNA. The MassArray instrument is able to quantitatively determine the proportion of methylated vs. unmethylated DNA. However, due to the reduced sequence complexity of bisulfite-converted DNA, the often-occurring overlap of peaks is difficult to interpret. Therefore, not all CpG sites can be analyzed for a given bisulfite-PCR product. These two methods are capable of measuring 96–384 PCR products in a single reaction and provide a medium-throughput platform for candidate gene methylation analysis.

3. Whole genome bisulfite sequencing (WGBS-seq)

Whole genome shotgun bisulfite sequencing (WGBS-seq) is the best way of determining the landscape of the whole methylome. WGBS-seq has been successfully used to map the complete methylomes of several human embryonic stem (ES) cell lines [34,35], human peripheral mononuclear cells [36], and hematopoietic progenitor cells at the single methylcytosine resolution [37]. The single-base level analysis was instrumental in the discovery of non-CpG methylation in ES cells [35]. Recently, the WGBS-seq analyses of several cancer samples have been completed [38–40]. Again, WGBS-seq made a significant contribution to the discovery of large partially methylated domains in cancer cells [38,40]. However, despite its advantages, the WGBS-seq remains too expensive to be applied to a large number of samples. In order to achieve higher sensitivity capable of detecting methylation differences between samples, a greater sequencing depth is required ultimately leading to a significant increase in the cost of sequencing. Currently, most of the WGBS-seq studies were conducted using Illumina GAII or HiSeq 2000 sequencers, while only one study used the SOLiD sequencer [40]. Although we have previously used 454-sequencing for genome-wide bisulfite sequencing of methylation enriched DNA [31], it is cost-prohibitive to use 454-sequencing for WGBS-seq.

4. Targeted bisulfite sequencing (TBS-seq)

The development of the targeted sequencing approach was mainly driven by the desire to conduct whole exome sequencing that would allow for the identification of disease-causing genes at a low cost compared to whole genome sequencing. A number of methods for targeted genome sequencing have been developed [41–44]. These approaches have been gradually integrated with bisulfite sequencing over time. Table 1 summarizes the characteristics and some of the key performance parameters of these TBS-seq approaches. The following section will explore each of these approaches in more detail.

4.1. Bisulfite padlock probes (BSPPs)

Padlock probes are single strand DNA fragments (100–150 nucleotide long) designed to hybridize to genomic DNA targets in a horseshoe manner (Fig. 1A) [45]. The targeted region for sequence capture is the gap between the two hybridized, locus-specific arms of a padlock probe. After the gap is filled in with a polymerase and ligated to form a circular strand of DNA, the target sequence is enriched by digestion of the linear DNA with nucleases. The target sequence in the circular DNA can be amplified using the common 'backbone' sequence that connects the two arms and is eventually converted into a sequencing library. This approach enables tens of thousands of probes to be used in a single reaction. Bisulfite padlock probes (BSPP) follow the same principle, but they are designed based on bisulfite-converted DNA sequences. Deng et al. used the BSPP approach to assess the methylation status of ~66,000 CpG sites in 2020 CGI on human chromosomes 12 and 20 [46]. Ball and colleagues designed ~10,000 padlock probes to profile ~7000 CpGs within the ENCODE pilot project regions [47]. Both studies demonstrated that the BSPP approach is highly

Please cite this article in press as: E.-J. Lee et al., Analyzing the cancer methylome through targeted bisulfite sequencing, Cancer Lett. (2012), http://dx.doi.org/10.1016/j.canlet.2012.10.040