Methods 47 (2009) 142–150

Contents lists available at ScienceDirect

Methods

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

Genome-wide high throughput analysis of DNA methylation in eukaryotes

Kyle R. Pomraning, Kristina M. Smith, Michael Freitag*

Center for Genome Research and Biocomputing and Department of Biochemistry and Biophysics, Oregon State University, 2011 ALS Building, Corvallis, OR 97331-7305, USA

ARTICLE INFO

Article history:
Accepted 29 September 2008
Available online 23 October 2008

Keywords:
Cytosine DNA methylation
High throughput sequencing
5-methylcytosine
MeDIP
Methylome
Epigenome
Epigenetics
Bisulfite sequencing
Genomic sequencing
Neurospora crassa

ABSTRACT

Cytosine methylation is the quintessential epigenetic mark. Two well-established methods, bisulfite sequencing and methyl-DNA immunoprecipitation (MeDIP) lend themselves to the genome-wide analysis of DNA methylation by high throughput sequencing. Here we provide an overview and brief review of these methods. We summarize our experience with MeDIP followed by high throughput Illumina/Solexa sequencing, exemplified by the analysis of the methylated fraction of the *Neurospora crassa* genome ("methylome"). We provide detailed methods for DNA isolation, processing and the generation of *in vitro* libraries for Illumina/Solexa sequencing. We discuss potential problems in the generation of sequencing libraries. Finally, we provide an overview of software that is appropriate for the analysis of high throughput sequencing data generated by Illumina/Solexa-type sequencing by synthesis, with a special emphasis on approaches and applications that can generate more accurate depictions of sequence reads that fall in repeated regions of a chosen reference genome.

1. Introduction

Site-specific DNA methylation is an epigenetic mark found in organisms across all domains of life. Archae and eubacteria have a variety of methylated nucleotides including *N*4-methylcytosine (4mC), 5-methylcytosine (5mC) and *N*6-methyladenine (6mA) [1,2]. The 6mA found at GATC sites in α-proteobacteria is likely the best studied DNA modification in single celled organisms and is involved in genome defense, DNA mismatch repair, DNA replication and control of gene expression [3]. While some eukaryotes (e.g. the genus *Tetrahymena*) contain 6mA, 5mC is the most prevalent, most studied and best understood DNA modification in eukaryotes [4]. Here we describe genome-wide approaches to study the distribution of 5mC in eukaryotes.

In many eukaryotes, DNA methylation is thought to control gene expression by modulation of DNA–protein interactions [5–7]. Nevertheless, DNA methylation is not essential in all eukaryotes. The yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* and the nematode *Caenorhabditis elegans* appear to have lost the DNA methylation machinery as no DNA methyltransferase genes are present in their genomes and no DNA methylation has been detected by various methods [8–10]. Very low levels of DNA methylation have been reported during development in *Drosophila melanogaster* [11]. Within the filamentous fungi, DNA methylation has been detected in many species but it is not universally present, e.g. *Aspergillus nidulans* has no detectable DNA methylation while very low levels have been reported from the closely related *A. flavus* [12]. Two fungi that have been used extensively for DNA methylation research, *Neurospora crassa* and *Ascobolus immersus*, show heterogeneous distribution of 5mC in all possible sequence contexts. In these species, DNA methylation is thought to be involved in "genome defense", e.g. by silencing invading transposable elements [13], but is not essential for survival. In plants, e.g. *Arabidopsis thaliana*, DNA methylation similarly silences the expression of transposable elements. Upon loss of DNA methylation, transposons reactivate and integrate in various regions of the genome, causing pleiotropic effects on development as has been demonstrated by use of *ddm1* mutants [14]. Arabidopsis contains an abundance of 5mC sites [15] in symmetrical and non-symmetrical contexts at CpG, CpHpG and CpHpH sites, where H represents A, C or T [16,17]. In vertebrates, where DNA methylation plays a role in chromatin packaging [18,19] and transcription [20–24], 5mC occurs almost exclusively in the CpG context [25].

Many methods have been used to assess DNA methylation [26]. HPLC is a conceptionally simple method to determine the overall content of 5mC in genomic DNA [27], and has been used in combination with thin-layer chromatography to demonstrate very minor amounts of DNA methylation in *D. melanogaster* and *A. flavus* [11,12]. Nevertheless, establishing absence or presence of DNA methylation and determining the percentage of 5mC in a specific genome is only the first step to elucidate methylation patterns that may control gene

* Corresponding author. Fax: +1 541 737 0481.
E-mail address: freitagm@cgrb.oregonstate.edu (M. Freitag).

1046-2023/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2008.09.022

expression. Current methods are either used for "typing" a specific known region that may or may not be methylated, or for "profiling", where parts or all of a given genome are investigated and no *a priori* knowledge of methylation status and/or sequence is required [26].

One of the earliest typing approaches relies on the comparison of digestion patterns caused by methylation-sensitive and methylation-insensitive isoschizomers of restriction endonucleases [28,29]. This approach has proven useful to establish methylation patterns for many vertebrate, plant and fungal genes or intergenic regions. In fungi, analysis with isoschizomers that require symmetrical recognition sites resulted in the underestimation of 5mC content [30]; later studies showed that 5mC can occur in non-symmetrical sites in Neurospora [31]. Other methylation typing approaches involve semi-quantitative and quantitative PCR methods that are useful for detection and quantification of methylation at certain loci by gene amplification after restriction digest [32,33]. In most cases, bisulfite conversion of 5mC is the most informative way to analyze gene- or region-specific DNA methylation patterns. Unmodified cytosines are converted to uracil (U), while 5mC remains unconverted. The converted DNA is amplified by PCR, thus introducing C to T changes in the PCR products, which are sequenced and compared to untreated DNA controls to yield the exact position of 5mC within a sequence [34]. The most common pitfalls of this method are related to incomplete C–U conversion and the inability to find appropriate primer pair combinations in species where 5mC is not found in symmetrical contexts. Many combinations of the techniques described above were applied before the widespread use of microarrays and are reviewed elsewhere [35,36].

Restriction landmark genomic scanning (RLGS) was the first truly genome-wide method available for analyzing methylation by comparing restriction fragments after end-labeling of digested DNA [37]. Another early genome-wide approach relied on purification of methylated DNA by column chromatography on resins that were coupled to the methyl-binding domain (MBD) of MeCP2 [13,38,39]. More recently, microarray technology has been developed as a powerful tool to analyze the methylation state of an organism [40–43]. Methylated DNA immunoprecipitation (MeDIP) [44] is used in conjunction with tiling microarrays for the evaluation of specific genetic loci and disease states [41,45–48]. Many microarray variants have also been used in conjunction with bisulfite conversion to analyze the methylation status of specific loci of interest during disease [22,49–52] or whole genomes [53,54]. Currently, tiling array design and production are the main drawbacks to this technique and are confounded by inaccurate hybridization signals. Moreover, in most genomes relatively heavily methylated regions of repeat DNA have not been sequenced or assembled and are thus lacking from current microarray designs.

As a result, microarray technology is being replaced by high throughput short read sequencing as the method of choice. Here we describe two techniques as carried out with the filamentous fungus *Neurospora crassa*. MeDIP sequencing (MeDIP-Seq) combines immunoprecipitation of 5mC residues with sequencing to create moderate resolution methylation profiles, while traditional bisulfite conversion is used with high throughput sequencing (bisulfite sequencing, BS-Seq; formerly called "genomic sequencing") to generate single base resolution methylation profiles (Fig. 1). These techniques can be used in tandem to provide independent methods for validation of whole genome methylation profiling.

2. DNA isolation and fragmentation

2.1. Introduction

Prior to methylation analysis genomic DNA must be purified and processed. Isolation of genomic DNA and separation of the methylated from the non-methylated component of the genome are the first critical steps in determining genome-wide methylation patterns. We found that commonly used methods for the isolation of genomic DNA prove successful for subsequent MeDIP, as well as validation by region-specific PCR or Southern blotting. This includes many commercially available kits for the isolation of plant DNA. Here we include a detailed description of a widely used method for the preparation of genomic DNA from filamentous fungi.

2.2. DNA isolation from filamentous fungi

We grow Neurospora as shaking cultures in 10 ml of Vogel's minimal medium [55] with 1.5% sucrose at 32 °C for three days in 125 ml Erlenmeyer flasks or 22 × 150 mm culture tubes with KimCaps. Tissue is harvested with wooden applicator sticks, briefly dried on paper towels, transferred to 15 ml polypropylene tubes and frozen at −20 °C or −80 °C. Solid tissue is lyophilized overnight and pulverized by vortexing with a metal spatula for ∼30 s. Samples are suspended in salt–detergent solution: Dissolve

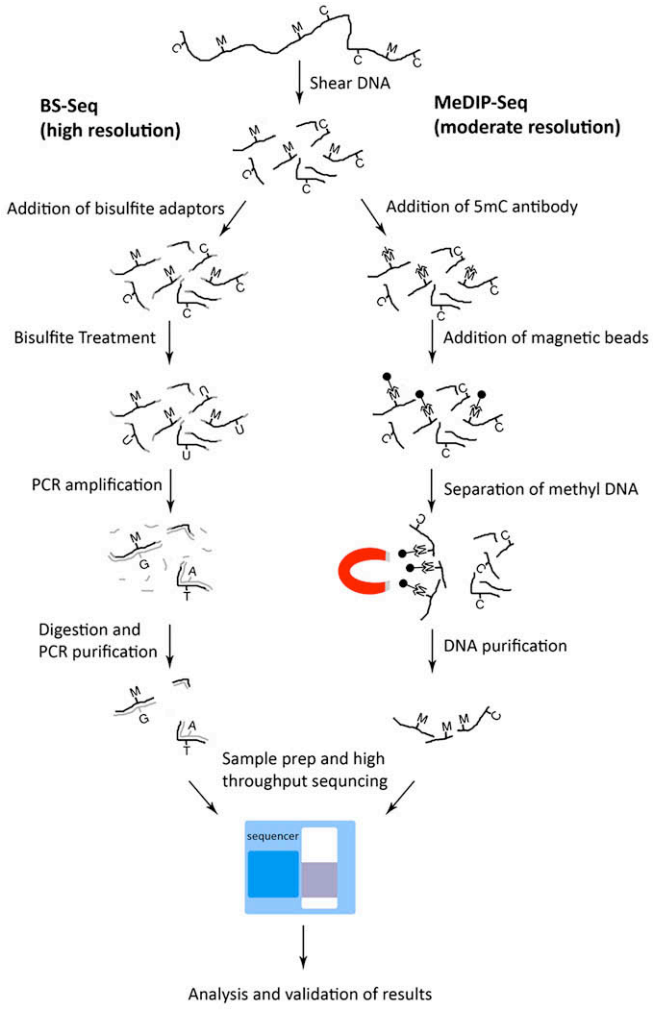

Fig. 1. Methods for whole genome methylation analysis. Genomic DNA is purified and sheared. For BS-Seq, (left pathway) adaptors are ligated to the DNA prior to treatment with bisulfite which converts unmethylated cytosines (C) into uracils (U). Methylated cytosines (M) remain unmodified. The converted DNA is amplified by PCR and digested. For MeDIP-Seq (right pathway), methylated DNA is immunoprecipitated with 5mC-specific antibodies. In both cases, the methylated fraction of the genome is purified, assembled into *in vitro* libraries and sequenced on high throughput sequencing machines. The resulting millions of short sequence "reads" are aligned to a reference genome to map enrichment of methylated DNA in specific chromosomal regions.

2 g of Na–deoxycholate (Sigma, D6750), 5 g of Brij 58/polyoxyethylene20cetyl ether (Sigma, P5884), 58.44 g of NaCl in 350 ml sterile distilled water, adjust to 500 ml with water and store at 4 °C;a precipitate may appear over time, but after re-mixing the solution can be used without ill effects. We add sufficient salt–detergent solution (usually ~600 μl) to the pulverized tissue to make a thick suspension after vortexing at high speed for 20 s. Samples are incubated at room temperature for ~20 min and vortexed once or twice. Tissues will become viscous over time. At this point samples can be transferred to Eppendorf tubes for further processing. Samples are centrifuged either in the large tube in a Sorvall SS-34 rotor at 8,000 rpm for 10 min, or at full speed in a 1.5 or 2 ml Eppendorf tube in a benchtop microcentrifuge. This typically yields 400–600 μl of supernatant, which is transferred to a fresh Eppendorf tube. The tube is filled with 4.5 M TCA:EtOH (1:1 v/v) solution. To make 4.5 M TCA:EtOH solution, dissolve 417 g Na-TCA salt (not free acid; Crescent Chemical Co., #AV17004) in 200 ml of water, use low heat to get the TCA salt into solution, adjust to 500 ml, add 500 ml of 95% or 100% ethanol and store at 4 °C. A precipitate will appear and settle, usually overnight—this precipitate should be avoided. At least 3 volumes of TCA:EtOH solution should be added for efficient precipitation. Samples are mixed by inversion and nucleic acids and proteins precipitated at −20 °C for at least 2 h. Samples may be stored at this stage for up to two days—upon further storage the resulting genomic DNA does not digest well with most restriction endonucleases. Nucleic acids are pelleted by centrifugation at full speed for 1 min in a microfuge and the supernatant aspirated or poured off the pellets. The pellets are resuspended in 200 μl of 10 mM NH_4OAc + 0.3 μg/ml RNase A (a pre-mixed solution that can be stored at 4 °C for several months), detached from tube walls by striking several times across a plastic tube rack and vortexed briefly. RNA is digested at 50 °C for 40 min; we vortex briefly every 10 min to help resuspend the pellet. We add 200 μl of chloroform, vortex briefly, centrifuge in a microfuge for 5 min at full speed and transfer the now clear supernatant to a fresh 1.5 ml tube. We add 900 μl of isopropanol/NH_4OAc (for a stock that can be kept at room temperature for several weeks, mix 42.5 ml isopropanol + 7.5 ml of 5 M NH_4OAc), mix well by inversion, centrifuge immediately in a microfuge for 1 min at full speed, aspirate the supernatant, wash the pellet with 300 μl of 70% EtOH, aspirate the supernatant and air-dry for 15 min. The pellet is resuspended in 100 μl TE buffer overnight at 4 °C, typically yielding ~200 ng/μl or ~25 μg of genomic DNA per culture.

During purification we include an RNase step, which reduces the size of contaminating RNA in the DNA preparation. An additional gel purification step after sonication but before MeDIP removes the remaining RNA and may be beneficial to increase recovery of DNA from the MeDIP, because the 5-methyl cytidine antibody used for MeDIP also recognizes methylated RNA.

2.3. DNA processing

Genomic DNA must be processed into small fragments prior to MeDIP-Seq or BS-Seq. Fragment ends need to be repaired and adaptors need to be ligated to the short fragments for both BS-Seq (see Section 4.2.) and high throughput sequencing methods (see Section 5.). Historically, cleavage into short fragments has been achieved using either restriction endonucleases, which generate non-random fragments based on DNA composition, or a variety of mechanical shearing techniques. Like others [44,56,57] (Z.A. Lewis, S. Honda, T. Khlafallah, J.K. Jeffress, M. Freitag, F. Mohn, D. Schübeler and E.U. Selker, submitted), we found that sonication is a relatively quick method to shear the DNA while avoiding the bias of restriction enzymes. End-repair and adaptor ligation protocols resemble those that have been used to generate probes for microarray analyses [56,58] For Neurospora, this MeDIP-Seq protocol works very well as there is sufficient time to allow reannealing of single stranded, immunoprecipitated DNA to generate double-stranded DNA, which serves as substrate for the polishing and adaptor ligation reactions. Alternatively, the polishing and adaptor ligation reactions are carried out prior to the immunoprecipitation steps [16,17].

To shear DNA, we dilute 5–50 μg of genomic DNA into 400 μl of TE buffer in a microcentrifuge tube. We have successfully used 5 μg of DNA to MeDIP the ~42 Mb Neurospora genome, where 5 μg represent ~100 million copies of the genome. For sonication we use a Branson sonifier 450 equipped with a microtip, set to a duty cycle of 80% and output control of 1.2 [59,60]. For other models specific sonication conditions need to be established on a case-by-case basis. We sonicate the DNA five times for 10 s each with 30 s rest on ice between cycles and run ~250 ng of sheared DNA on a 1% agarose gel to verify the size distribution of the sonicated DNA. We typically aim for a smear of sheared DNA between 300 and 1000 bp in length. If the smear of DNA is too large one can sonicate the remaining sample for additional cycles and re-check the fragment size distribution. Prior to working with a new organism we recommend running a standard from two to twelve cycles of sonication to determine how many cycles are necessary to achieve the desired result. With Neurospora DNA, even just two cycles of sonication reduces the average size of the genomic DNA to ~1.2 kb.

The size of the sheared DNA directly affects the precision achievable when mapping MeDIP-Seq reads. Because the 5-methyl cytidine antibody requires more than just a single 5mC to efficiently bind [63], shorter DNA fragments will allow more precise mapping to a reference genome but may not enrich for regions that have only a few methylated cytosines. Conversely, longer fragments will make mapping less precise but will increase sensitivity for detection of regions with lower levels of 5mC. The size of the sheared fragments will also affect PCR validation of the results as small DNA fragments will require closely spaced PCR primer pairs. Like others [44] and (Z.A. Lewis, S. Honda, T. Khlafallah, J.K. Jeffress, M. Freitag, F. Mohn, D. Schübeler and E.U. Selker, submitted), we found fragments sheared to ~500 bp to be the most useful. This allows us to design PCR primer pairs that amplify control products of ~400 bp along with test products of ~200 bp (see Section 7). One method to increase the accuracy associated with read mapping from high throughput sequencing is to decrease the size range of the input DNA by excising a narrow, well-defined band (450–550 bp), purifying the DNA with a commercially available gel extraction kit and using this as the input DNA for MeDIP.

3. MeDIP protocol

Our protocol is almost identical to the original MeDIP method described by Weber et al. [44] and was first used with Neurospora by Selker and co-workers (Z.A. Lewis, S. Honda, T. Khlafallah, J.K. Jeffress, M. Freitag, F. Mohn, D. Schübeler and E.U. Selker, submitted). We save a quarter of the sonicated input DNA as control and use the rest for the MeDIP. At least 5 μg of sheared DNA is diluted into 450 μl of TE buffer, denatured in a 100 °C heat block for 10 min and snap-cooled on ice for 5 min. We add 50 μl of 10× immunoprecipitation buffer [100 mM Na-Phosphate pH 7.0, 1.4 M NaCl, 0.5% TritonX-100] and 1 μl of the 5mC antibody to the DNA solution (Diagenode, #MAb-5MECYT-100, 1 μg/μl), and incubate for 2 h on an orbital rotator at 4 °C. While the antibodies and DNA are incubating, we pre-wash 40 μl of magnetic Dynabeads (Invitrogen, M-280 sheep anti-mouse IgG) with 1 ml of PBS + 0.1% BSA for 5 min at room temperature with shaking. The beads are collected by use of a stick magnet attached to a pipet tip rack, and

the wash is repeated once. Beads are resuspended in 40 μl of 1× immunoprecipitation buffer, added to the DNA sample and incubated on an end-over-end rotator at 4 °C for 2–16 h. Dynabeads are collected with a stick magnet as above and the supernatant with unbound DNA is removed. Beads are washed three times with 1 ml of 1× immunoprecipitation buffer for 10 min at room temperature with shaking, resuspended in 250 μl proteinase K digestion buffer [5 mM Tris pH 8.0, 1 mM EDTA pH 8.0, 0.05% SDS] with 7 μl of 10 mg/ml proteinase K and incubated for 3 h on an end-over-end rotator at 50 °C to digest the antibodies and release the 5mC-containing DNA. DNA is extracted once with 250 μl phenol, once with 250 μl chloroform and precipitated by adding 500 μl ethanol with 400 mM NaCl. To improve recovery, 1 μl glycogen (20 mg/ml) is added. DNA pellets are resuspended in 50 μl TE buffer and stored at −20 °C.

A convenient alternative to precipitation with 5mC antibodies is the use of commercially available MeDIP kits that rely on the interaction of the methyl-binding domain (MBD) of MBD2 or MeCP2 with 5mC [61,62]. While MeCP2 has a natural histidine (His) tag because its protein sequence in mice, rats and humans contains seven consecutive histidine residues [62,63], the MBD of MBD2 has been expressed with a His tag to allow purification with magnetic nickel beads (ActiveMotif, Carlsbad, CA). These types of methylated-CpG island recovery assays ("MIRA") have been used in several studies [64]. Different MBDs have slightly different affinities for the density and number of consecutive CpGs (i.e., for the MBD2 interaction to be efficient, several CpGs need to be in close proximity, whereas fewer, more widely spaced CpGs are sufficient for interaction with MeCP2 [65].

4. Bisulfite sequencing

Bisulfite (or "genomic") sequencing is useful to determine the methylation status of cytosines at the single nucleotide level. Briefly, single-stranded DNA is treated with bisulfite which sulfanates cytosine but leaves 5mC unaffected. The cytosine is then deaminated and desulfanated to uracil [66]. Converted DNA is amplified by PCR with convenient primer pairs and PCR products are directly sequenced and aligned to unconverted DNA, thus revealing exactly which cytosines were methylated. Two parameters affect the success of this technique: (1) complete conversion of unmethylated cytosines to uracil and (2) potential degradation of DNA during the conversion reaction by high temperature and low pH. Incomplete conversion mainly occurs because bisulfite only attacks cytosines in single-stranded DNA. In areas of the genome with high GC content the DNA may not denature completely, which results in patches of unmodified cytosines [67]. If a reasonably complete genome sequence is available, these false-positive regions can be screened for in an organism-specific manner. In a study to define the methylated fraction of the Arabidopsis genome (the "methylome"), a high proportion of false-positive sequences were removed from further analysis by screening for reads with three methylated CpHpH sites in a row [16].

4.1. End-repair of sheared DNA for bisulfite sequencing

Ligation of the sheared DNA to adaptors prior to bisulfite treatment should ensure unbiased PCR amplification of completely bisulfite-converted DNA. The end-repair of sheared DNA follows essentially the same protocol as used for the preparation of microarray probes [56] or Illumina/Solexa sequencing libraries (see details in Section 5). Polished DNA is mixed with Klenow polymerase (exo$^-$) and dATP to generate a 3′ A-overhang, purified on MinElute PCR purification columns (Qiagen, Valencia, CA) and eluted with 10 μl of the supplied elution buffer.

4.2. Design of PCR primers and adaptors for bisulfite conversion

A linker ligation step coupled to PCR after bisulfite treatment selects for sequences that have undergone complete cytosine to uracil conversion; because only 18 PCR cycles are used, this allows for unbiased amplification [68]. Linker sequences that contain unmethylated cytosines and a restriction site are annealed to template (Fig. 2). After bisulfite treatment, primers designed to amplify only completely bisulfite-converted linker sequences are used for PCR amplification. The restriction site (in Fig. 2 it is *Alu*I) should be as close to the template sequence as possible to minimize the amount of linker sequenced. To ligate double-stranded bisulfite sequencing adaptors, mix the adaptors and polished DNA (10:1 molar ratio), add 5 μl T4 DNA ligase (1 U/μl), 25 μl DNA ligase buffer and water to bring the reaction volume to 50 μl. Different commercial DNA ligase preparations result in similar efficiencies. Incubate the reaction at room temperature for 15 min. To remove unligated adaptors, separate the ligated DNA on a 1% agarose gel, exise the 300–500 bp fraction and extract the DNA on commercially available PCR purification column (Qiagen; Valencia, CA; Epoch Biolabs. Houston, TX); elute with 32 μl of elution buffer (10 mM Tris–HCl, pH 8.5).

4.3. Bisulfite conversion of DNA

By now, a number of kits for bisulfite sequencing are commercially available, all of which promise to minimize degradation while maximizing the C to U conversion rate. Recently, the "CpGenome DNA modification kit" (Chemicon; Temecula, CA) and the "EpiTect kit" (Qiagen; Valencia, CA) were both shown to produce C to U conversion levels of greater than 99% with only modest DNA degradation [16,17]. Other techniques have been shown to achieve even more complete conversion rates but at the expense of DNA integrity [17,66]. After bisulfite conversion, bisulfite-converted DNA is amplified by 18 PCR cycles with primers that are specific for bisulfite sequencing adaptors (Fig. 2). The DNA is digested with the restriction enzyme whose recognition site was included on the adaptors, purified on commercially available PCR purification columns (Qiagen; Valencia, CA; Epoch Biolabs. Houston, TX) and used to generate *in vitro* sequencing libraries.

Fig. 2. Preparation of sheared DNA for bisulfite sequencing. (a) Double-stranded adaptor sequences (black) are ligated to sheared end-repaired DNA (red). The DNA is treated with bisulfite, which converts unmethylated cytosine to uracil. (b) A linker-specific PCR primer (blue) is used to selectively amplify top-strand sequences that have undergone complete bisulfite conversion. (c) In the second round of PCR, a second conversion-specific primer amplifies bottom-strand sequences. (d) After 18 rounds of PCR, the products will consist primarily of bisulfite-converted DNA in which all U/G mismatches have been converted to T/A. The adaptors are removed at the indicated sites by digestion with restriction endonucleases, e.g. *Alu*I.

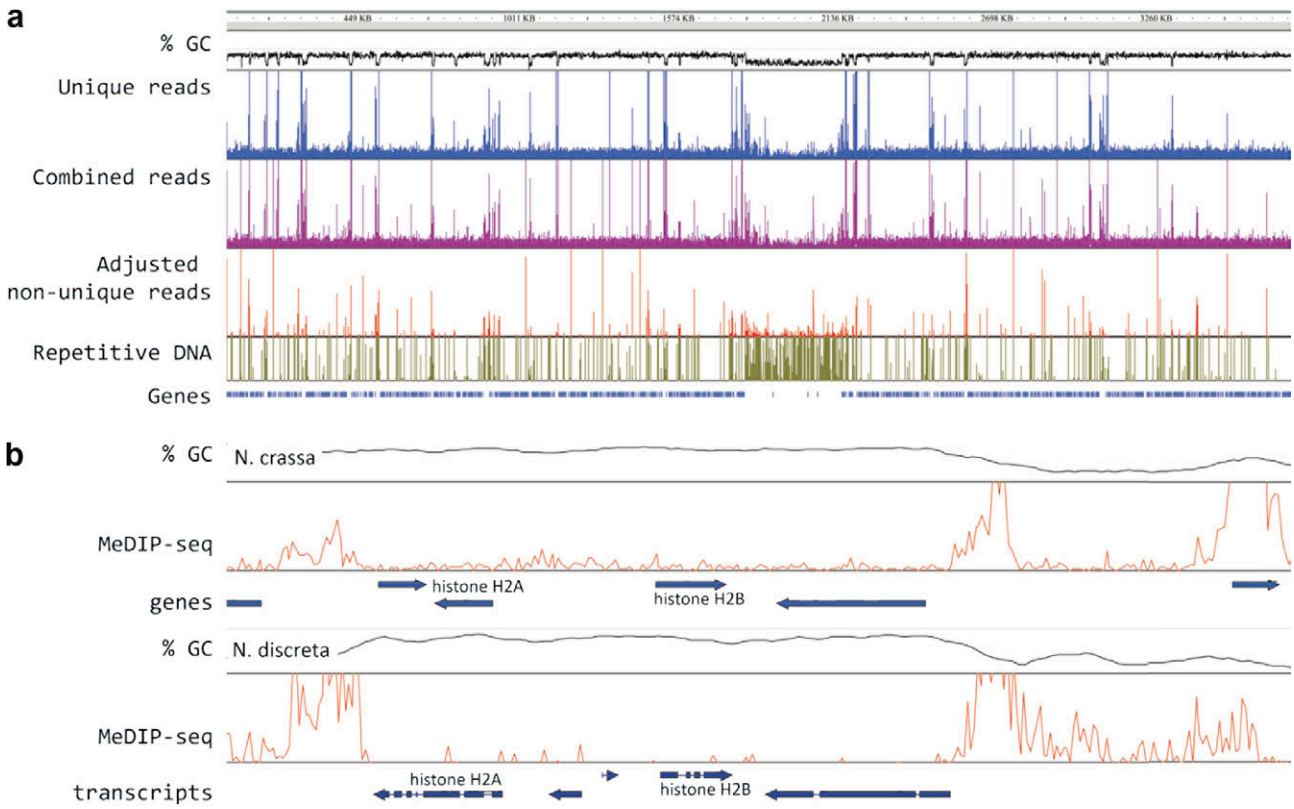

Fig. 3. DNA methylation of linkage group (LG) VII in *Neurospora crassa*. (a) Unique (blue) and non-unique reads (red) were aligned to the current *N. crassa* reference sequence for LG VII by BLAT. Methylation peaks were visualized in the Argo browser as histograms in 50 bp windows, generated by our "BLATmapper" script. Repetitive DNA (gold) was mapped by counting how many times each possible 32-mer occurs in the Neurospora genome. This track provides a visual guide for repetitiveness of non-unique read mapping. Read map height was determined by counting the number of times a read was sequenced and dividing by the number of times that read occurs in the genome. (b) DNA methylation peaks (red) in a syntenic region that flanks the LG VII centromeres of *N. crassa* (top) and the related species *N. discreta* (bottom). DNA methylation in both species occurs almost exclusively in intergenic regions. The ORF for the heavily methylated "predicted protein" on the right of the *N. crassa* track appears to be a pseudogene.

5. Preparation of DNA libraries for Illumina/Solexa sequencing

A number of high throughput sequencing technologies are currently available, including pyrosequencing (Roche/454), sequencing by ligation (ABI SOLiD), and reversible terminator sequencing (Illumina/Solexa). These technologies give the ability to quickly and inexpensively sequence very large amounts of DNA but all have the drawback of generating short (~400 nt) or very short (36–50 nt) sequence reads. We make use of Solexa sequencing, which is currently able to generate ~1.5 Gb of sequence data in ~3 days from a single flow cell with eight channels.

Prior to Solexa sequencing, DNA samples from MeDIP or bisulfite conversions must be processed to generate *in vitro* sequencing libraries. Illumina/Solexa recommends use of their genomic or chromatin immunoprecipitation (ChIP) sample preparation kits. Nevertheless, except for the adaptor and PCR primers the reagents supplied with the kit are not in any way different from the typically used enzymes available in most molecular biology labs. Illumina supplies modified primers as separate "primer-only" kits (and the type and extent of modification are considered proprietary information). We found that unmodified HPLC-purified or non-purified primers can work as well as Illumina-supplied primer kits; this reduces the sample preparation costs by about threefold.

5.1. End-repair of sheared DNA for Illumina/Solexa sequencing

When sequencing MeDIP samples, we typically start with 200–400 ng of DNA. Sheared DNA fragments are repaired with T4 DNA polymerase to fill in 5′ overhangs, Klenow polymerase to remove the 3′ overhangs, and T4 polynucleotide kinase to phosphorylate the 5′-OH. As suggested in the Illumina sample preparation protocols, we mix 30 μl of the DNA to be sequenced with 45 μl H_2O, 10 μl of 10× T4 DNA ligase buffer with 10 mM ATP, 2 μl dNTP mix (20 mM of each dNTP), 5 μl T4 DNA polymerase (3 U/μl), 1 μl Klenow polymerase (5 U/μl) and 5 μl T4 polynucleotide kinase (10 U/μl). In lieu of Ilumina sample preparation kits, enzymes from other manufacturers can be substituted with good success when used at the concentrations given above. The polishing reaction proceeds at 20 °C for 30 min. Polished DNA is purified on commercially available purification columns (e.g. Qiagen, Valencia, CA; Epoch Biolabs, Houston, TX), and eluted with 32 μl of 10 mM Tris–HCl (pH 8.5). To allow ligation to the Illumina/Solexa adaptor primers, an adenine needs to be added to the 3′ ends of DNA fragments by using Klenow polymerase that lacks the 3′ to 5′ exonuclease activity. The polished DNA (32 μl) is mixed with 5 μl of 10× Klenow polymerase buffer, 10 μl of 1 mM ATP and 3 μl Klenow (exo^-) polymerase (5 U/μl). The reaction is incubated at 37 °C for 30 min. We purify the DNA on MinElute PCR purification columns (Qiagen, Valencia, CA) and elute with 10 μl of 10 mM Tris–HCl (pH 8.5).

5.2. Adaptor ligation and PCR amplification for Illumina/Solexa sequencing

The adaptor primers provided by Solexa are denatured by heating to 98 °C for 5 min and cooled to room temperature. Alternatively, equal amounts of lab-designed primers (at 10 μM) are mixed, denatured by heating to 98 °C for 5 min and cooled to room temperature. Annealed adaptors are stable at −20 °C for at least

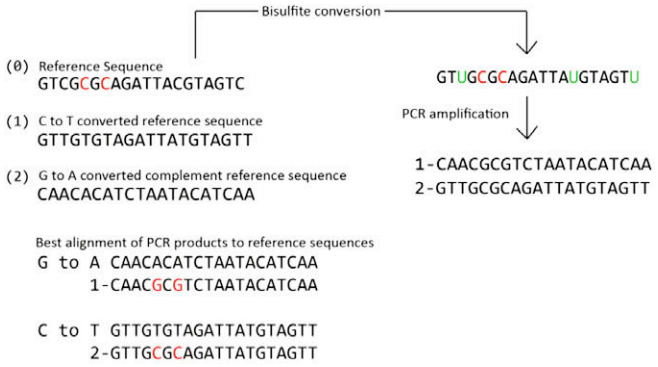

Fig. 4. Analysis of reads from high throughput bisulfite sequencing. To map sequence reads derived from BS-Seq, two additional reference genomes are prepared from a current reference genome (0). The first (1) is the reference genome with all cytosines changed to thymines. The second (2) is the complement sequence to the genome with all guanines changed to adenines. After bisulfite conversion the DNA is subjected to PCR amplification resulting in two main products from any given sequence. The products are sequenced and aligned to their best hits in either of the two converted reference genomes. C/T mismatches (in the C to T converted reference sequence) and G/A mismatches (in the G to A converted complement reference sequence) indicate the position of a methylated cytosine. Methylated cytosines are red while uracils derived from converted unmethylated cytosines are shown in green.

one year. Annealed adaptors are ligated to the DNA by mixing them with polished DNA in a molar ratio of 10:1. Mix the adaptors, DNA, 5 μl T4 DNA ligase (1 U/μl), ligase buffer, and water (final volume of 50 μl). Incubate at room temperature for 15 min. Separate ligated DNA from unligated adapters on a 2% NuSieve agarose gel and select a DNA size for Solexa sequencing. We typically excise DNA between 200 and 500 bp. We purify this DNA with a commercially available gel extraction kit (e.g. Qiagen or Epoch Biolabs). The *in vitro* library is amplified by PCR and thus enriched for DNA fragments that are flanked by ligated adaptors. We use 1 μl of the gel-purified DNA with 25 μl Phusion DNA polymerase Master Mix (Finnzymes, NEB), 1 μl PCR primer 1.1 (Illumina), 1 μl PCR primer 2.1 (Illumina) and 22 μl of water. After initial denaturation (30 s at 98 °C), only 18–24 cycles of PCR are used (10 s at 98 °C, 30 s at 65 °C, 30 s at 72 °C) to avoid selective amplification of specific regions. This is followed by a 5 min extension at 72 °C. Amplicons are purified on PCR purification columns and eluted with 30–50 μl of 10 mM Tris–HCl (pH 8.5). DNA concentration is measured by absorbance at 260 nm (e.g. on a Nanodrop spectrophotometer) and ~10% of the reaction is separated on a 1–2% agarose gel to monitor amplification.

Sequencing libraries are diluted to the specifications required by the Illumina/Solexa sequencing machine in use. In most situations, cluster generation and the actual sequencing is handled by personnel who are part of a core sequencing facility. As these manipulations do not lend themselves to changes or specific adaptations by individual labs, these steps are not discussed here any further. Manuals and protocols that describe these steps are available from Illumina/Solexa.

6. Data analysis and visualization

6.1. Introduction

High throughput sequencing, e.g. with the Illumina/Solexa 1G genome analyzer, generates nearly a terabyte of image files during a single run. Image files are analyzed and converted into sequence reads. Prior to undertaking any high throughput sequencing project it is essential to design and test a data analysis pipeline. Dealing with the large amount of data that is generated from even a single run is not a trivial task and in most lab environments this will require investment in additional computing resources. Additionally, it is recommended that genome browsers (e.g. Gmod/GBrowse or Argo) are installed and functional to allow visualization of mapped sequence reads in a user-friendly manner.

Until recently, the read length of Illumina/Solexa sequencing was 36 bp. Because of a drastic increase of sequencing errors at the end of the reads, most mapping programs only consider the first 32 bp; usable reads have been significantly increased with the release of the new version of the Illumina/Solexa sequencer. After a complete list of reads for all eight lanes or channels on a single flowcell is compiled, individual reads are mapped onto the reference genome. For that purpose, Illumina provides the ELAND algorithm, which is able to efficiently map reads of 32 bp to a reference genome. This program yields six datasets. The first three datasets are comprised of reads that map to a unique location in the reference genome with zero, one, or two mismatches (U0, U1, U2, respectively). The second three datasets contain reads that map to more than one place in the reference genome with zero, one, or two mismatches (R0, R1, R2, respectively). In many cases, unique reads are used for most if not all of the data analyses, as ELAND provides map coordinates for the unique data sets. A drawback to the ELAND algorithm is that it does not give coordinates for non-unique reads. In order to overcome this, non-unique hits must be mapped back to the genome using a program such as BLAT [69]; traditional BLAST algorithms are not recommended for the analysis of high throughput data.

Efficient alternative methods are now available to map non-unique reads. Two of these programs have been developed at Oregon State University, CashX (J. Cumbie, C. Sullivan, K. Kasschau and J. Carrington, in preparation) and RGA ("reference-guided assembly"; R. Shen and T. Mockler, in preparation). We use both programs, along with BLAT, to map repetitive DNA sequences that are methylated. CashX uses tabulated and tallied reads to map them precisely to the reference genome while keeping track of read tally and all map positions in the genomes. Output data are transferred into files that can be used as track input for commonly used genome browsers (e.g. in GFF3 format). RGA uses a BLAT-type mapping approach, resulting in a simple table or histogram format that can be plotted with several programs, e.g. "R". SOAP ("small oligo alignment program") is another useful mapping program for the visualization of Illumina/Solexa data [70]. In all three applications, a variable number of mismatches and gaps are allowed to facilitate mapping.

No matter how non-unique reads are mapped back to a reference genome, it is critical to represent data in an unbiased manner and validate calls for non-unique reads. Various normalization methods have been proposed [71]. The problem lies in not knowing precisely where a specific non-unique read originated. Several ways to solve this conundrum can be imagined. Reads can be mapped to all possible locations and "normalized" by simply dividing by the number of times a read hits to the genome. To make this approach more sophisticated, reads can be analyzed by nearest-neighbor approaches to evaluate if one specific region in the genome is more likely than others to have produced a certain percentage of non-unique reads. For example, if only one out of five potential regions shows unique reads around a small number of non-unique reads, it is more plausible that this single region originated non-unique reads. Dividing the number of non-unique reads by the number of genomic locations would generate a read count for this special region that would be too low.

Other than already published normalization approaches, assignment of a confidence value to each hit provides one method for evaluating the mapping of non-unique sequences (Fig. 3a). This approach can be used prior to validation of mapping data by other methods (e.g. region-specific quantitative PCR). We assign a "read value" (i.e., number of times a particular oligonucleotide is sequenced divided by the number of places it maps in the genome) and a "confidence

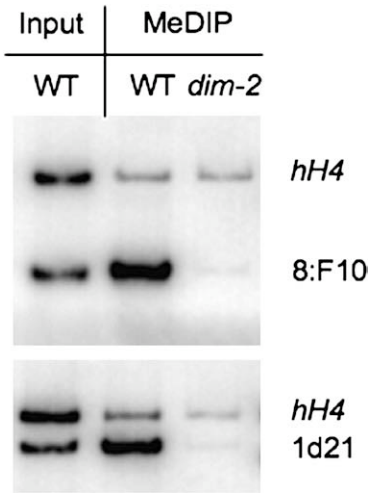

Fig. 5. Validation of DNA methylation by semi-quantitative region-specific PCR. Primers were designed for two regions known to be methylated in Neurospora (8:F10 and 1d21) and for an unmethylated control region, the histone H4 gene (*hH4*). To quantify the amount of DNA methylation in these regions, the ratio of the band intensities between the MeDIP and control samples are calculated. In the wildtype (wt) MeDIP both methylated regions showed enrichment when compared to the sonicated wt control. In contrast, a DNA methyltransferase mutant (*dim-2*) that lacks all DNA methylation known in *N. crassa* showed no enrichment in the MeDIP sample.

value" (i.e., 1/number of places a read maps to the genome). Thus, a read that maps to a unique location in the genome will be given a confidence value of 1, while a read that maps to 60 different places will be given a confidence value of 1/60 (=0.0167). To further facilitate viewing, the read value can be used to assign the height in a histogram while the confidence value can be simultaneously mapped as a heat map (e.g. a read with a high confidence value is mapped as red while a low confidence value is mapped as blue on a spectrum).

All mapping approaches require the use of genome browsers to efficiently view mapped reads in relation to other genome annotations. In most situations, this requires investment in computational resources to allow the use of Gmod/Gbrowse, which needs to be set up specifically for each organism that is to be analyzed [72]. The most current information on this open source tool is available at http://gmod.org/wiki/index.php/Gbrowse. Other browsers can fulfill useful roles for specific requirements. To view the relatively small (4–11 Mb) chromosomes of filamentous fungi in combination with MeDIP-seq or ChIP-seq data tracks we frequently use Argo (Available at http://www.broad.mit.edu/annotation/argo/) because Gbrowse appears too cumbersome to display large segments of a genome in a single screenshot.

6.2. MeDIP sequencing

Currently, reads generated during a Illumina/Solexa sequencing run are 36 bp in length but represent fragments of DNA that were anywhere between 200 and 500 bp in length depending on the size selected when gel purifying the adapter-ligated library. Each start position of a read mapped to the genome represents the 5′ end of one of these sheared pieces. Therefore, each read that is aligned to the genome comes with some amount of uncertainty as to the exact location of the methylated cytosines. For example, if fragments of ~200–500 bp were selected, in theory the methylated cytosines are located anywhere within a 1000 bp window with the read start position at the center. With good coverage, reads will stack to yield a good prediction of the methylated regions as shown in Fig. 3. Methylated regions from two closely related *Neurospora* species, *N. crassa* and *N. discreta* coincide almost perfectly in some regions, e.g. the segment of linkage group VII around the single histone H2A and histone H2B genes (Fig. 3b).

6.3. Bisulfite sequencing

Aligning bisulfite-converted reads to a reference genome is essentially the same as aligning non-converted reads with the caveat of not knowing whether a thymine base call is actually a thymine or a bisulfite-converted cytosine. However, if high conversion rates of umethylated cytosines were observed, >99% of the remaining cytosines can be considered to be methylated. Additionally, since the complementary strand of the DNA will also be altered during PCR amplification it is not known whether an adenine is actually an adenine or whether it was the complement guanine to a converted cytosine. A straight forward method to overcome this problem was used by Lister and co-workers to map methylation in the Arabidopsis genome [17] while a similar method is described in detail by Cokus and co-workers [16].

Three reference genomes must be used for alignment of bisulfite-converted reads (Fig. 4). The first is an unconverted genome. The second is a genome where all the cytosines are converted to thymines to represent complete conversion of unmethylated cytosines and the third is a complement genome where all the reference genome has been complemented and the guanines are converted to adenines to represent PCR conversion of the complementary DNA. Methylated cytosines can be identified when a read aligns to a converted genome with a mismatch. If the mismatched base is a cytosine aligning to a thymine or a guanine aligning to an adenine and the position was originally a cytosine or guanine in the unconverted genome then the mismatch represents a methylated cytosine.

When mapping reads to the converted genomes it is important to keep in mind that reads with higher concentrations of methylated cytosines will map poorly. For example, the ELAND mapping program normally allows up to two mismatches during alignment, which will occur when mapping a methylated cytosine onto the converted genomes. However, if an area contains very high levels of methylation (greater than two methylated cytosines in a 36 bp read) it will not map to the converted genome but may map better to the unconverted genome if there are fewer than two unmethylated cytosines in the bisulfite-converted read.

7. Validation of results

Validation of results is crucial for high throughput applications and provides an estimate of the data's accuracy. Quantitative and semi-quantitative PCR are straightforward methods for validation of methylation levels at specific loci. In either case, the level of a genomic sequence of interest and a control sequence are compared between control and MeDIP samples with enrichment in the MeDIP sample indicating cytosine methylation. Southern analysis may also be used to look for the presence of a specific DNA fragment. We also have mapped 5mC in the related organisms *N. crassa* and *N. discreta* (Fig. 3b) and compared similar regions of their genomes to verify broad methylation patterns.

We use primarily semi-quantitative PCR with radioactive labeling for validation of MeDIP and ChIP-sequencing (Fig. 5). Real-time PCR is a convenient alternative if available. We use primers for a known euchromatic sequence without DNA methylation (e.g. part of the Neurospora histone H4 gene, *hH4*) as a negative control. DNA sequences known to be methylated serve as positive controls. Several primer sets for sequences of interest are designed. Typically we design the negative and positive control product to be ~400 bp and the test product to be ~200 bp. This allows co-amplification in one PCR reaction and facilitates separation and quantification after running PAGE gels and exposing phosphorimager cassettes or film to the dried gels.

For semi-quantitive PCR, mix 12.65 μl H2O, 0.5 μl polymerase (2 U/μl), 0.25 μl dNTP mix (20 mM), 2.5 μl 10× polymerase buffer, 0.1 μl α-^{32}P labeled dCTP (20 mM), 0.5 μl template DNA, 2 μl control primers (5 mM each), 2 μl test primers (5 mM each) and amplify the DNA by PCR (90 s at 94 °C, 24 times 30 s at 94 °C, 30 s at 55 °C, 30 s at 72 °C, and 5 min at 72 °C)[60]. We separate 10 μl of each PCR product by PAGE through a 5% gel, record the band intensities by phosphorimaging and calculate the relative intensity of the test band divided by the control band for each sample. The relative intensities from the MeDIP and control samples are compared to find enrichment in the MeDIP sample (Fig. 5).

8. Concluding remarks

High throughput sequencing has revolutionized the analysis of methylated DNA. In non-repeat regions, 5mC can now be mapped with exquisite precision by genome-wide bisulfite sequencing. The relatively short read lengths generated by most high throughput sequencing approaches continue to stifle attempts to provide the complete "methylome" of most higher eukaryotes, simply because much 5mC resides in regions that consist of repeated DNA. Nevertheless, both MeDIP-Seq and BS-Seq have now been used to map methylation patterns in the non-repetitive regions of fungal, plant and animal genomes. Genetic studies combined with high throughput sequencing are carried out both in wildtype and mutant strains. In mutants the capacity to methylate DNA can either be grossly or only slightly altered, thus revealing differential methylation patterns that help to uncover the mechanisms of DNA methylation. These and future studies will provide valuable insights into how eukaryotic genomes are organized, how gene clusters or regulons are coordinately regulated by epigenetic mechanisms, and how DNA methylation affects the expression of specific genes.

Acknowledgments

Our laboratory is supported by a Grant from the American Cancer Society (RSG-08-030-01-CCG, to M.F.). We are indebted to Eric Selker for the original Neurospora MeDIP protocol. We thank Zack Lewis and Eric Selker for sharing of unpublished data, materials and stimulating discussions. High throughput sequencing and initial data analyses were carried out in the OSU CGRB core lab by Mark Dasenko, Steve Drake and Chris Sullivan. We are grateful for gifts of strains and DNA oligomers from the Fungal Genetics Stock Center (FGSC, University of Kansas, Kansas City, MS) and the Neurospora Genome Program Project grant from the NIH (P01 GM068087), respectively.

References

[1] M. Ehrlich, M.A. Gama-Sosa, L.H. Carreira, L.G. Ljungdahl, K.C. Kuo, C.W. Gehrke, Nucleic Acids Res. 13 (1985) 1399–1412.
[2] D. Lodwick, H.N. Ross, J.E. Harris, J.W. Almond, W.D. Grant, J. Gen. Microbiol. 132 (1986) 3055–3059.
[3] M.G. Marinus, Annu. Rev. Genet. 21 (1987) 113–131.
[4] D. Ratel, J.L. Ravanat, F. Berger, D. Wion, Bioessays 28 (2006) 309–315.
[5] J.D. Lewis, R.R. Meehan, W.J. Henzel, I. Maurer-Fogy, P. Jeppesen, F. Klein, A. Bird, Cell 69 (1992) 905–914.
[6] J.H. Lee, D.G. Skalnik, J. Biol. Chem. 277 (2002) 42259–42267.
[7] P.H. Tate, A.P. Bird, Curr. Opin. Genet. Dev. 3 (1993) 226–231.
[8] J.H. Proffitt, J.R. Davie, D. Swinton, S. Hattman, Mol. Cell Biol. 4 (1984) 985–988.
[9] V.J. Simpson, T.E. Johnson, R.F. Hammen, Nucleic Acids Res. 14 (1986) 6711–6719.
[10] C.R. Wilkinson, R. Bartlett, P. Nurse, A.P. Bird, Nucleic Acids Res. 23 (1995) 203–210.
[11] H. Gowher, O. Leismann, A. Jeltsch, EMBO J. 19 (2000) 6918–6923.
[12] H. Gowher, K.C. Ehrlich, A. Jeltsch, FEMS Microbiol. Lett. 205 (2001) 151–155.
[13] E.U. Selker, N.A. Tountas, S.H. Cross, B.S. Margolin, J.G. Murphy, A.P. Bird, M. Freitag, Nature 422 (2003) 893–897.
[14] A. Miura, S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada, T. Kakutani, Nature 411 (2001) 212–214.
[15] R.E. Pruitt, E.M. Meyerowitz, J. Mol. Biol. 187 (1986) 169–183.
[16] S.J. Cokus, S. Feng, X. Zhang, Z. Chen, B. Merriman, C.D. Haudenschild, S. Pradhan, S.F. Nelson, M. Pellegrini, S.E. Jacobsen, Nature 452 (2008) 215–219.
[17] R. Lister, R.C. O'Malley, J. Tonti-Filippini, B.D. Gregory, C.C. Berry, A.H. Millar, J.R. Ecker, Cell 133 (2008) 523–536.
[18] M. Weber, I. Hellmann, M.B. Stadler, L. Ramos, S. Paabo, M. Rebhan, D. Schubeler, Nat. Genet. 39 (2007) 457–466.
[19] J. Lewis, A. Bird, FEBS Lett. 285 (1991) 155–159.
[20] R. Metivier, R. Gallais, C. Tiffoche, C. Le Peron, R.Z. Jurkowska, R.P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, Nature 452 (2008) 45–50.
[21] G. Gokul, B. Gautami, S. Malathi, A.P. Sowjanya, U.R. Poli, M. Jain, G. Ramakrishna, S. Khosla, Epigenetics 2 (2007) 80–85.
[22] Z. Zhang, P.C. Huettner, L. Nguyen, M. Bidder, M.C. Funk, J. Li, J.S. Rader, Oncogene 25 (2006) 5436–5445.
[23] Z. Wang, Y. Zhang, B. Ramsahoye, D. Bowen, S.H. Lim, Br. J. Cancer. 91 (2004) 1597–1603.
[24] N. Allaman-Pillet, A. Djemai, C. Bonny, D.F. Schorderet, Gene Expr. 7 (1998) 61–73.
[25] M.J. Browne, R.H. Burdon, Nucleic Acids Res. 4 (1977) 1025–1037.
[26] S. Beck, V.K. Rakyan, Trends Genet. 24 (2008) 231–237.
[27] K.C. Kuo, R.A. McCune, C.W. Gehrke, R. Midgett, M. Ehrlich, Nucleic Acids Res. 8 (1980) 4763–4776.
[28] M. Nelson, E. Raschke, M. McClelland, Nucl. Acids Res. 21 (1993) 3139–3154.
[29] A.P. Bird, E.M. Southern, J. Mol. Biol. 118 (1978) 27–47.
[30] F. Antequera, M. Tamame, J.R. Villanueva, T. Santos, J. Biol. Chem. 259 (1984) 8033–8036.
[31] E.U. Selker, J.N. Stevens, Proc, Natl. Acad. Sci. USA 82 (1985) 8114–8118.
[32] M. Heiskanen, A.C. Syvanen, H. Siitari, S. Laine, A. Palotie, PCR Methods Appl. 4 (1994) 26–30.
[33] J. Singer-Sam, J.M. LeBon, R.L. Tanguay, A.D. Riggs, Nucleic Acids Res. 18 (1990) 687.
[34] M. Frommer, L.E. McDonald, D.S. Millar, C.M. Collis, F. Watt, G.W. Grigg, P.L. Molloy, C.L. Paul, Proc. Natl. Acad. Sci. USA 89 (1992) 1827–1831.
[35] S. Derks, M.H. Lentjes, D.M. Hellebrekers, A.P. de Bruine, J.G. Herman, M. van Engeland, Cell Oncol. 26 (2004) 291–299.
[36] M.F. Fraga, M. Esteller, Biotechniques 33 (2002) 636–649.
[37] H. Cedar, A. Solage, G. Glaser, A. Razin, Nucleic Acids Res. 6 (1979) 2125–2132.
[38] S.H. Cross, J.A. Charlton, X. Nan, A.P. Bird, Nat. Genet. 6 (1994) 236–244.
[39] M. Shiraishi, A. Sekiguchi, Y.H. Chuu, T. Sekiya, Biol. Chem. 380 (1999) 1127–1131.
[40] M.R. Estecio, P.S. Yan, A.E. Ibrahim, C.S. Tellez, L. Shen, T.H. Huang, J.P. Issa, Genome Res. 17 (2007) 1529–1536.
[41] E. Schilling, M. Rehli, Genomics 90 (2007) 314–323.
[42] P.S. Yan, C.M. Chen, H. Shi, F. Rahmatpanah, S.H. Wei, C.W. Caldwell, T.H. Huang, Cancer Res. 61 (2001) 8375–8380.
[43] S.E. Brown, M.F. Fraga, I.C. Weaver, M. Berdasco, M. Szyf, Epigenetics 2 (2007) 54–65.
[44] M. Weber, J.J. Davies, D. Wittig, E.J. Oakeley, M. Haase, W.L. Lam, D. Schubeler, Nat. Genet. 37 (2005) 853–862.
[45] A.S. Cheng, A.C. Culhane, M.W. Chan, C.R. Venkataramu, M. Ehrich, A. Nasir, B.A. Rodriguez, J. Liu, P.S. Yan, J. Quackenbush, K.P. Nephew, T.J. Yeatman, T.H. Huang, Cancer Res. 68 (2008) 1786–1796.
[46] X.Q. Tian, D.F. Sun, Y.J. Zhang, J.Y. Fang, Yi Chuan 30 (2008) 295–303.
[47] F.V. Jacinto, E. Ballestar, S. Ropero, M. Esteller, Cancer Res. 67 (2007) 11481–11486.
[48] C. Gebhard, L. Schwarzfischer, T.H. Pham, E. Schilling, M. Klug, R. Andreesen, M. Rehli, Cancer Res. 66 (2006) 6118–6128.
[49] K. So, G. Tamura, T. Honda, N. Homma, T. Waki, N. Togawa, S. Nishizuka, T. Motoyama, Cancer Sci. 97 (2006) 1155–1158.
[50] N. Omura, C.P. Li, A. Li, S.M. Hong, K. Walter, A. Jimeno, M. Hidalgo, M. Goggins, Cancer Biol. Ther. 7 (2008) .
[51] M.O. Hoque, M.S. Kim, K.L. Ostrow, J. Liu, G.B. Wisman, H.L. Park, M.L. Poeta, C. Jeronimo, R. Henrique, A. Lendvai, E. Schuuring, S. Begum, E. Rosenbaum, M. Ongenaert, K. Yamashita, J. Califano, W. Westra, A.G. van der Zee, W. Van Criekinge, D. Sidransky, Cancer Res. 68 (2008) 2661–2670.
[52] D. Zhou, W. Qiao, Y. Wan, Z. Lu, J. Biochem. Biophys. Methods 66 (2006) 33–43.
[53] J. Penterman, D. Zilberman, J.H. Huh, T. Ballinger, S. Henikoff, R.L. Fischer, Proc. Natl. Acad. Sci. USA 104 (2007) 6752–6757.
[54] H. Hayashi, G. Nagae, S. Tsutsumi, K. Kaneshiro, T. Kozaki, A. Kaneda, H. Sugisaki, H. Aburatani, Hum. Genet. 120 (2007) 701–711.
[55] 333 R.H. Davis, Oxford University Press, 2000 333.
[56] M.R. Miller, T.S. Atwood, B.F. Eames, J.K. Eberhart, Y.L. Yan, J.H. Postlethwait, E.A. Johnson, Genome Biol. 8 (2007) R105.
[57] Z.A. Lewis, A.L. Shiver, N. Stiffler, M.R. Miller, E.A. Johnson, E.U. Selker, Genetics 177 (2007) 1163–1171.
[58] J.L. DeRisi, V.R. Iyer, P.O. Brown, Science 278 (1997) 680–686.
[59] J. Nakayama, A.J. Klar, S.I. Grewal, Cell 101 (2000) 307–317.
[60] H. Tamaru, X. Zhang, D. McMillen, P.B. Singh, J. Nakayama, S.I. Grewal, C.D. Allis, X. Cheng, E.U. Selker, Nat. Genet. 34 (2003) 75–79.
[61] B. Hendrich, A. Bird, Mol. Cellular Biol. 18 (1998) 6538–6547.
[62] X. Nan, R.R. Meehan, A. Bird, Nucleic Acids Res. 21 (1993) 4886–4892.
[63] R.R. Meehan, J.D. Lewis, A.P. Bird, Nucleic Acids Res. 20 (1992) 5085–5092.
[64] T. Rauch, H. Li, X. Wu, G.P. Pfeifer, Cancer Res. 66 (2006) 7939–7947.
[65] M.F. Fraga, E. Ballestar, G. Montoya, P. Taysavang, P.A. Wade, M. Esteller, Nucleic Acids Res. 31 (2003) 1765–1774.

[66] S.J. Clark, A. Statham, C. Stirzaker, P.L. Molloy, M. Frommer, Nat. Protoc. 1 (2006) 2353–2364.
[67] P.M. Warnecke, C. Stirzaker, J. Song, C. Grunau, J.R. Melki, S.J. Clark, Methods 27 (2002) 101–107.
[68] A. Meissner, A. Gnirke, G.W. Bell, B. Ramsahoye, E.S. Lander, R. Jaenisch, Nucleic Acids Res. 33 (2005) 5868–5877.
[69] W.J. Kent, Genome Res. 12 (2002) 656–664.
[70] R. Li, Y. Li, K. Kristiansen, J. Wang, Bioinformatics 24 (2008) 713–714.
[71] K.D. Kasschau, N. Fahlgren, E.J. Chapman, C.M. Sullivan, J.S. Cumbie, S.A. Givan, J.C. Carrington, PLoS Biol. 5 (2007) e57.
[72] M.J. Donlin, Curr Protoc Bioinformatics Chapter 9 (2007) Unit 9 9.