Chapter 8

Bisulfite Sequencing of Cloned Alleles

Zhiqing Huang, Christopher F. Bassil, and Susan K. Murphy

Abstract

Bisulfite sequencing of cloned alleles is a widely used method for capturing the methylation profiles of single alleles. This method combines PCR amplification of the bisulfite-modified DNA with the subcloning of the amplicons into plasmids followed by transformation into bacteria and plating on selective media. The resulting colony forming units are each comprised of bacterial clones containing the same plasmid reflecting a single allele in the original PCR reaction. Following whole cell PCR and sequencing, the results provide highly detailed information about the status of each CG site within an allele. Sequencing of a large number of individual clones can provide quantitative information, assuming unbiased PCR, subcloning and clone selection. The proportion of methylated cytosine at a particular position within the sequenced alleles can be determined by counting the number of alleles showing methylation at the position of interest and dividing this by the total number of clones sequenced.

Key words DNA methylation, Sodium bisulfite, Cloning, Capillary sequencing

1 Introduction

The cloning and sequencing of bisulfite converted DNA is one of the gold standards of DNA methylation analysis, as the sequencing of subcloned individual DNA molecules provides detailed information on the methylation status of every CG site within a given allele for relatively long stretches of DNA sequence [1–3]. The genomic DNA is first modified by sodium bisulfite. The DNA sequence under investigation is amplified by PCR with primers specific for one strand of the bisulfite converted DNA. The amplified PCR product comprises a pool of DNA molecules representing the individual alleles present within the DNA specimen used for PCR. A portion of the PCR product is subcloned into plasmids and randomly selected clones are sequenced to provide the methylation status of the CG sites within each of the individual DNA molecules. Methylation percentages for each of the CG sites are often calculated according to the number of methylated and unmethylated CG cytosines at a given position across the different clones

Anastasia Malek and Oleg Tchernitsa (eds.), *Ovarian Cancer: Methods and Protocols*, Methods in Molecular Biology, vol. 1049,
DOI 10.1007/978-1-62703-547-7_8,

that are sequenced. Accuracy increases with the number of clones sequenced.

Relative to direct sequencing of a PCR amplicon pool generated from a bisulfite-modified genomic DNA template, sequencing of clones provides a higher resolution result due to the ability to score methylation for single alleles. The sequence data is semi-quantitative and can reveal variable patterns of methylation within a specimen. However, the procedure can be labor intensive. In addition, PCR bias and/or cloning bias may contribute to skewing of results when attempting to quantify methylation [4].

We begin with important considerations for primer design. Subheading 3 is divided into subsections that describe PCR amplification and purification, ligation, transformation and screening, sequencing of the cloned alleles and basic interpretation of results.

To avoid preferential amplification of methylated versus umethylated DNA, the primers for bisulfite sequencing must be designed to anneal to regions devoid of CG dinucleotides. In addition, when selecting primers, a database with information about the presence of polymorphisms (GeneCards, dbSNP, SNP500, SNPbrowser) should be consulted to avoid designing primers that anneal to potential polymorphic sites. Primers for bisulfite sequencing have only three bases and are T-rich due to conversion of non-CG cytosines. This often requires that the primers be greater than 30 bases in length to obtain primer melting temperatures of ~60 °C.

Primer annealing sites can be difficult to identify due to limited availability of regions of sufficient length within an otherwise CG-rich stretch of sequence without inclusion of one or more potentially methylated CG dinucleotide(s). If there are no suitable regions for primer design, the antisense strand derived from the reverse complement of the original input genomic DNA sequence may offer a suitable alternative. The complementary nature of the sense and antisense strands is eliminated following bisulfite modification. However, the same number of CGs and their methylation state (due to the palindromic nature of the methylation status of CG sequence motifs) will be present in both the sense and antisense strands and thus either strand can be used to define the methylation profile of a particular region.

Controls (e.g., no template and unmodified genomic DNA) should be included when running the PCR reactions to insure that there is no contamination and that the amplification is specific to the bisulfite converted DNA.

Bisulfite sequencing is still widely used in research labs, largely because it is straightforward, the primer design is relatively simple, and it provides highly detailed methylation information for the region of interest without the need for specialized instrumentation. This technique is especially useful for studies in which the particular pattern of methylation is of interest, such as studies of site-specific methylation, or for study of allele-specific methylation

as is characteristic of the methylation patterns at the regulatory regions associated with genomically imprinted genes, or genes subject to X chromosome inactivation. Limitations of this technique include that it can be time-consuming and expensive depending on the desired number of clones to be sequenced and the number of specimens to analyze. In addition, the accuracy of results obtained can be adversely affected by bias in PCR amplification efficiency and bias in subcloning for alleles of a particular methylation status. Long stretches of thymine nucleotides that are common in bisulfite-modified DNA can also cause problems with the accuracy of the sequencing [5]. Nevertheless, this is the only available technique aside from high throughput bisulfite sequencing approaches that provides allele-specific methylation status at single nucleotide resolution, and thus promises to be a mainstay for analysis of DNA methylation.

2 Materials

2.1 Kits

1. HotStarTaq Polymerase kit (Qiagen, Valencia, USA) or Platinum Taq PCR kit (Invitrogen, Grand Island, USA), containing 10× PCR buffer, 25 mM $MgCl_2$ stock, DNA polymerase and nuclease-free water.
2. Qiagen MinElute PCR Purification kit (Qiagen, Valencia, USA).
3. BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, USA).
4. pGEM®-T Easy Vector System II (containing JM109 competent cells) (Promega, Madison, USA).
5. Epitect Unmethylated and Methylated Bisulfite Converted Controls (Qiagen, Valencia, USA).

2.2 Reagent and Supplies

1. Bisulfite converted DNA (20 ng/μl). The bisulfite converted DNA is stable for up to 6 months at −80 °C, up to 3 months at −20 °C and for 1–2 weeks at 4 °C. Avoid freeze-thaws.
2. Forward and revers primers for PCR amplification of target sequence (*see* **Note 1**).
3. T7 and SP6 primers for whole cell PCR amplification.
4. 100 mM dNTP set, PCR grade (Invitrogen, Carlsbad, USA). dNTPs should be stored at −20 °C in 10–20 μl aliquots. Dilute to 10 mM with Nuclease-free water before use.
5. GenePure LE Quick Dissolve Agarose (ISC BioExpress, Kaysville, USA).
6. Parafilm® M Barrier Film (SPI Supplies, West Chester, USA).
7. Ethidium bromide in dropper bottle (0.625 mg/mL) (Genesee Scientific, Research Triangle Park, USA).

8. Sigma GenElute spin columns (Sigma-Aldrich Corp. St. Louis, USA).
9. Phenol:Chloroform:Isoamyl Alcohol solution 25:24:1, Store at 4 °C.
10. 3 M sodium acetate (pH 5.5) (Ambion by Life Technologies, Grand Island, USA).
11. Tris-Borate-EDTA buffer (TBE): 0.089 M Tris-base, 0.089 M Boric acid, 0.002 M EDTA, pH 8.3.
12. 10× OrangeG Gel Loading Buffer. Dissolve 15 mL of glycerol and 100 mg of Orange G powder in water up to final volume 50 mL, vortex to mix. Store at 4 °C.
13. 50 bp DNA Ladder, 1,000 μg/mL (New England Biolabs, Ipswich, USA). Prior to use, add 20 μL ladder to 160 μL nuclease-free water. Add 20 μL OrangeG Gel Loading Buffer. Store at 4 °C.
14. LB agar with 100 μg/mL ampicillin and 60 μg/mL X-gal (Teknova Science Matters, Hollister, USA). Store at 4 °C and protect from light.
15. SOC medium (Sigma-Aldrich Corp. St. Louis, USA). Store at 4°C.

2.3 Equipment

1. Thermocycler.
2. Gel casting tray and gel running boxes for 8 × 10 cm gels.
3. Power supply.
4. UV transilluminator.
5. High throughput capillary sequencer (e.g., Applied Biosytems 3730 ×l) and associated software.

3 Methods

3.1 PCR and PCR Products Purification

This protocol is based on using the pGEM®-T Easy Vector System II, HotStarTaq PCR kit and the Qiagen MinElute PCR Purification kit. Use of any other reagents will require optimization.

3.1.1 PCR

Amplify bisulfite-modified genomic DNA (20–50 ng) using the HotStarTaq PCR kit and primers specific to the bisulfite-modified DNA according to the manufacturer's protocol.

1. Prepare the PCR Master Mix. An example for a 25 μL PCR reaction using the Qiagen HotStar Taq PCR kit is shown in Table 1 (*see* **Note 2**):
2. Thoroughly mix the PCR Master Mix by vortexing for several seconds. Spin down briefly to insure liquid is collected at the bottom of the tube.

Table 1
PCR master mix

Stock	Reaction mix volumes (μL)	Final concentration
10× PCR Buffer	2.5	1×
25 mM $MgCl_2$	1.5	1.5 mM
dNTPs (10 mM each)	0.5	0.2 mM each
Forward primer	1	100–500 nM
Reverse primer	1	100–500 nM
HotStarTaq DNA polymerase	0.2	2.0–2.5 U
DNA (to be added last)	1	10–50 ng
Nuclease-free water	17.3	
Total volume	25	

3. Transfer the appropriate amount of PCR Master Mix (25 μl minus the volume of DNA to be added) to reaction wells (*see* **Note 3**).
4. Add template DNA to Master Mix in reaction wells (*see* **Note 4**).
5. Seal the reaction wells. Mix the PCR mixture by vortexing for several seconds. Spin down briefly.
6. Transfer reaction wells containing PCR mixture to a thermocycler. Perform PCR as follows: Initialization step 95°C/15′, then cycling by denaturation step 94°C/30″, annealing step *X*°C/30″ (*see* **Note 5**), extension step 72°C/30″, for 25–35 cycles followed by extension at 72°C for 10′. The samples can be analyzed immediately or can be stored at –20°C prior to use.

3.1.2 Gel Electrophoresis

1. For an 8 × 10 cm, 2% agarose gel, add 50 mL of 1× TBE Buffer to 1.0 g of agarose in a glass beaker (*see* **Note 6**). Swirl gently to mix.
2. Microwave the buffer/gel mixture until the agarose is fully dissolved (usually 1–2 min). Watch closely, as it may be necessary to pause the microwave occasionally for a few seconds to avoid having the mixture boil over.
3. In a fume hood, add one drop of ethidium bromide solution (*see* **Note 7**) to the molten agarose mixture (final concentration 0.5 μg/mL) and swirl gently until fully mixed.
4. Still in the fume hood, carefully pour molten agarose into a gel casting tray with barrier ends and comb(s) in place. Let solidify for 15–20 min.

5. Remove the end barriers from the gel. Fill the gel box with 1× TBE buffer so that the surface of the gel is just submerged in the buffer solution. Carefully remove the comb(s) using a very gentle rocking motion while constantly applying slight upward pressure.
6. Add 10 μL of prepared 50 bp DNA ladder to the desired marker lane(s) for each row of samples to be run.
7. For each of the samples that will be run on the gel, pipet 1 μL of Orange G Gel Loading Buffer onto a short strip of Parafilm® M Barrier Film. The liquid will bead up on the film, making it easy to pipet. Carefully pipet 10 μL of the PCR product onto one of the Orange G drops, pipet up and down to mix, then slowly load the sample into the corresponding well of the gel, making sure to keep the pipet tip within the well while slowly dispensing the sample so that any run over of sample into adjacent wells is prevented. Repeat for the remaining samples.
8. Run gel at an appropriate voltage for the desired time (*see* **Note 8**).

3.1.3 Purification of PCR Products

1. Resolve PCR amplicons on a 2% agarose gel stained with ethidium bromide, as described above (*see* **Note 9**).
2. Excise the band containing the amplicon of interest from the gel (*see* **Note 10**), place the excised gel fragment into a GenElute spin column placed in one of the provided microcentrifuge tubes.
3. Spin in a balanced microcentrifuge at full speed for 15 min. Agarose will be retained on the filter in the spin column while the DNA will pass through the filter with the gel buffer and collect in the microcentrifuge tube. Discard the spin column.
4. Purify the eluted DNA using a standard phenol:chloroform: isoamyl alcohol extraction protocol followed by ethanol precipitation. Briefly, add an equal volume of phenol: chloroform:isoamyl alcohol solution to the DNA from **step 3** under Subheading 3.1.3, Vortex for 10 s followed by centrifugation for 15 min at full speed. Transfer the top (aqueous) layer into a clean tube. Add 1/10 volume 3 M sodium acetate (pH 5.5) and 2.5 volumes of absolute ethanol. Mix by briefly vortexing. Centrifuge for 30 min at full speed at 4°C. Carefully aspirate the ethanol and allow the DNA pellet to air dry. Resuspend the DNA in 10 μl of nuclease-free water.

3.2 Ligation, Transformation and Screening

The presented ligation and cloning protocols are adopted for the pGEM®-T Easy Vector System II, while whole-cell PCR is performed using the Platinum Taq PCR kit. Use of any other reagents will require optimization.

1. Ligate the purified PCR amplicons (1–3 μl) into pGEM®-T Easy plasmids, following the protocol from the manufacturer.
2. Allow the ligation to proceed for 8 h at 16-18°C, or overnight at 4°C.
3. Place 5–10 μl of each ligation reaction into a sterile 1.5 mL microcentrifuge tube on ice.
4. Remove the frozen JM109 High Efficiency Competent Cells from –80°C storage and place in an ice bath until just thawed (about 5 min). Mix the cells by gently flicking the tube.
5. Carefully transfer 50 μl of cells into each tube containing the ligation reaction. Gently flick the tubes to mix and place them on ice for 30 min.
6. Heat-shock the cells for 45–50 s (up to 2 min) in a water bath or heating block at exactly 42°C (do not shake).
7. Immediately return the tubes to ice for 2 min.
8. Add 1,000 μl room temperature SOC medium to the tubes containing the cells transformed with the ligation reactions (*see* **Note 11**)
9. Incubate for 1 h at 37°C with shaking (~150 rpm).
10. Using sterile technique, spread 100–500 μl of each transformation mixture onto LB/ampicillin/X-Gal plates (*see* **Note 12**).
11. Incubate the plates overnight (16–24 h) at 37°C. White colonies generally contain inserts.
12. Prepare reaction mix for appropriate number of PCRs using the Platinum Taq PCR kit according to the manufacturer's instructions.
13. Use whole-cell PCR to amplify the inserted DNA sequence within the plasmids from individual colonies with SP6 and T7 primers (binding sites for these primers are present within the plasmid, flanking the insertion site). The colonies can be picked from the plate using a sterile P-10 pipet tip and immediately transferred into the PCR reaction mix (*see* **Note 13**). Rotate the pipet tip within and against the sides of the reaction well while the tip is submerged in the liquid to transfer as many of the bacterial cells as possible into the PCR reaction mix. Pipet up and down to mix.
14. Perform PCR as follows: Initialization at 94°C for 5′, then 35 cycles of denaturation (94°C/30″), annealing (55°C/30″) and extension (72°C/30-45″), followed by a final extension at 72°C for 5′. Please see comments regarding PCR conditions in **Notes 14** and **15**.
15. Resolve amplicons on 2% agarose gels.
16. Calculate the expected size of the amplicons, including the relevant portions of the plasmid backbone (i.e., all of the T7

Table 2
Sequencing reaction mix (BigDye® Terminator v3.1 Cycle Sequencing Kit)

Stock	Reaction volumes, each
5× BigDye sequencing buffer	2 μL
Primer (10 μM)	1 μL
BigDye Terminator v3.1	1 μL
Eluted PCR amplicons	1–2 μL
Nuclease free water	x
Total volume	10 μL

primer binding site to the insertion site, the insert itself, and other side of the insertion site through the SP6 primer binding site).

17. Purify individual amplicons of the anticipated size using Sigma GenElute spin columns as described above, **Steps 2** and **3** of Subsection 3.1.3.

3.3 Sequencing of Cloned Alleles

The eluted amplicons are directly sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit from Applied Biosystems that utilizes the process of "dye terminator sequencing". The forward or reverse PCR primer used to amplify the region of interest (from Section 3.1.1 above) can generally be used for the sequencing reaction. The reactions are run using a capillary sequencing instrument, such as an ABI-3730xl-DNA-Analyzer. Results can be visualized with 4Peaks software (freeware from mekentosj.com) or Finch TV (free for download from PerkinElmer).

1. Prepare sequencing reaction mix according to the supplier's protocol. An example of preparing the reaction mix using the BigDye® Terminator v3.1 Cycle Sequencing Kit is shown in Table 2.
2. Amplify according to the manufacturer's recommendations: Initialization step 96°C/2′, then cycling by denaturation step 96°C/30″, annealing step 50°C/15″, extension step 62°C/240″, for 25–35 cycles.
3. Sequencing is performed in accordance with protocols provided by the equipment supplier.

3.4 Interpreting Results

The sequence obtained for each cloned allele should be inspected to insure that the sequence is clean and specific to the region of interest (*see* **Note 16**). Each clone showing good quality sequence can then be scored for methylation status at each individual CG dinucleotide. Using this technique, methylation status within a clone at a given CG position is a binary variable. Each CG cytosine is either methylated or unmethylated within the given allele.

The combined methylation status for each clone by CG site can be assembled and represented as shown in Fig. 1, where each row corresponds to one of the individual cloned alleles, and each column represents an individual CG site. Typically in such a

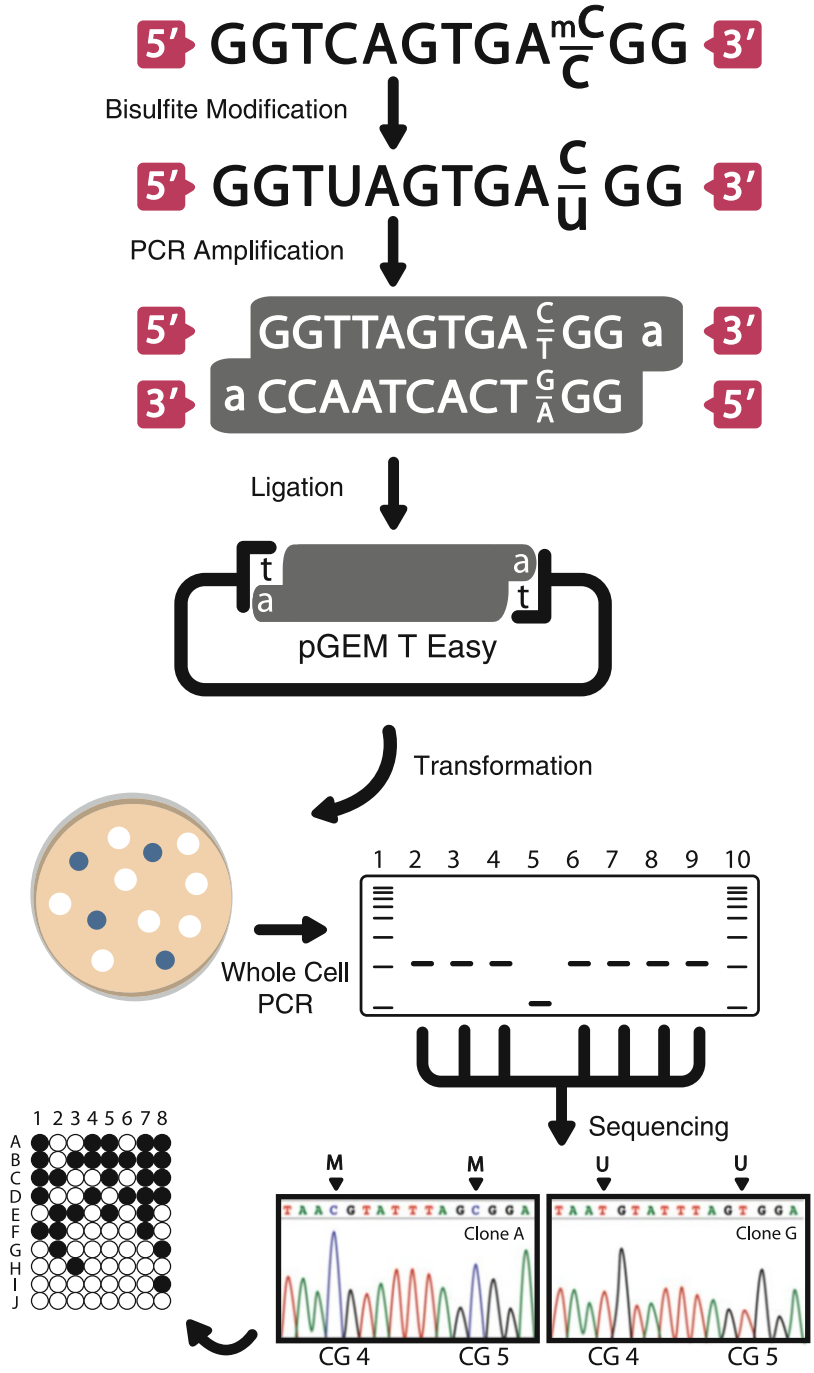

Fig. 1 (continued)

lollipop-style map, filled symbols are indicative of a methylated cytosine while unfilled symbols represent unmethylated cytosines. The methylation status of CG cytosines at a given position within the sequence, given as percent methylation, can be calculated as follows: $\%M = [C^M/(C^M + C^U)] \times 100$, where C^M is the number of clones with methylation at the particular cytosine being analyzed, and C^U is the number of clones unmethylated at the particular cytosine being analyzed. Accuracy is improved by increasing the number of clones analyzed, provided there is not bias in the cloning or amplification steps.

4 Notes

1. It is desirable to design primers such that they will anneal to sequence that includes several non-CG cytosines (thymines after bisulfite modification and PCR amplification) in the primer sequence to increase specificity for genomic DNA that has undergone complete conversion by the sodium bisulfite. High Performance Liquid Chromatography (HPLC) purification of the primers helps to assure that the primers are full length, thus improving the specificity of the primers for the intended target sequence.
2. Optimal reaction conditions (incubation times and temperatures, concentrations of primers, $MgCl_2$, and template DNA) vary and need to be optimized.
3. Controls (e.g., no template and unmodified genomic DNA) should be included when running the PCR reactions to insure that there is no contamination and that the amplification is specific to the bisulfite converted DNA.

Fig. 1 Bisulfite sequencing of cloned alleles. Bisulfite modification of genomic DNA is followed by PCR amplification and ligation of the amplicons into a plasmid backbone. The ligated DNA is then transformed into competent bacterial cells and allowed to grow briefly in non-selective liquid culture; the transformed cells are then plated to selective media and grown overnight. Blue/white screening allows for enhanced selection of (white) colony forming units that contain the PCR amplicons as inserts. Whole cell PCR followed by agarose gel electrophoresis is used to screen the selected colony forming units for the appropriately sized inserts. Positive amplicons are purified and sequenced using fluorescent BigDye sequencing (schematic of results shown). Each one of the colony forming units selected for sequencing is a clonal expansion of one bacterial cell containing a single allele from the original pool of PCR amplicons. The CG dinculeotides within the sequenced region are individually scored as "methylated" (e.g., CG sites 4 and 5 of Clone A) or "unmethylated" (e.g., CG sites 4 and 5 of Clone G). This type of data is often represented as a series of circles, each representing a single CG site, and arranged as shown in the figure. The individual rows represent one of the cloned alleles sequenced while the columns represent each of the CG dinucleotides within the sequenced region in the order they appear in the sequence, regardless of spacing between the CG sites within the original sequence. Filled circles represent methylated CG sites and unfilled circles represent unmethylated CG sites

4. The template DNA required varies and should be optimized. We typically elute 800 ng of bisulfite converted DNA with 40 μL of elution buffer (with the concentration of bisulfite converted DNA at about 20 ng/μL, assuming that all DNA was recovered after the bisulfite modification process) and use 1–3 μL (20–60 ng of input DNA) in each 25 μL PCR reaction.
5. The primer annealing temperatures are a good point at which to start when optimizing the PCR reactions, using the lowest annealing temperature for the primers used. Every effort should be made to try and choose primers with similar annealing temperatures to avoid bias in performance during the PCR reactions.
6. The concentration of TBE used to prepare the gel must match that used as running buffer. If the gel is prepared with 1X TBE, it must be run in 1X TBE running buffer.
7. Ethidium bromide intercalates double stranded DNA and is therefore a suspected mutagen. However, it is not considered hazardous waste at low concentrations. Guidelines for disposal vary. Those relevant to your institution or place of employment should be followed.
8. The timing of the run and power used will vary depending on the size of the gel, the percentage of agarose used to prepare the gel and the distance required for the amplicons to be resolved by migration through the gel.
9. Strong and specific PCR amplicons should be obtained from the bisulfite-modified genomic DNA in order to ensure cloning and sequencing success.
10. Protect your eyes and skin from UV exposure by wearing an appropriate UV protective face shield, long sleeved lab coat and gloves.
11. LB broth may be substituted, but colony number may be lower.
12. High colony density may make selecting individual CFUs difficult; it is better to have fewer CFUs such that each colony represents a clonal expansion of a single bacterial cell.
13. Usually a minimum of 10–20 white colony forming units (CFUs) are selected for analysis.
14. Denaturation at 96°C for 5 min before PCR thermocycling is necessary when performing whole cell PCR to facilitate release of plasmid DNA from the bacterial cells.
15. The extension time for the PCR reaction is size dependent. We typically use 30 s for an amplicon size less than 500 bp and 45 s for amplicons longer than 500 bp.

16. High quality sequence data from BigDye® sequencing typically begins approximately 50–60 bases from the 3′-end of the sequencing primer. The positioning of the sequencing primer should therefore be taken into account with regard to the sequence information desired.

Acknowledgments

We gratefully acknowledge Zack Davenport for his contributions to the artwork. We thank Allison Barratt and Carole Grenier for critical reading of the manuscript. This work was supported by Department of Defense grant W81XWH-11-1-0469, NIH grants R01DK085173 and R01CA142983 and the Gail Parkins Ovarian Cancer Awareness Fund.

References

1. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89:1827–1831
2. Clark SJ, Harrison J, Paul CL, Frommer M (1994) High sensitivity mapping of methylated cytosines. Nucleic Acids Res 22:2990–2997
3. Feil R, Charlton J, Bird AP, Walter J, Reik W (1994) Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. Nucleic Acids Res 22: 695–696
4. Warnecke PM, Stirzaker C, Song J, Grunau C, Melki JR et al (2002) Identification and resolution of artifacts in bisulfite sequencing. Methods 27:101–107
5. Kristensen LS, Hansen LL (2009) PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. Clin Chem 55:1471–1483