Title

Efficient and fast identification of differentially methylated regions using whole-genome bisulfite sequencing data

Journal

Journal of Genetics and Genomics, 45(8)

ISSN

1673-8527

Authors

Diep, Dinh

Zhang, Kun

Publication Date

2018-08-01

DOI

10.1016/j.jgg.2018.07.008

Peer reviewed

Chapter 19

Large-Scale Targeted DNA Methylation Analysis Using Bisulfite Padlock Probes

Dinh Diep, Nongluk Plongthongkum, and Kun Zhang

Abstract

Bisulfite padlock probes (BSPP) are a method for the targeted quantification of DNA methylation in mammalian genomes. They can simultaneously characterize the level of methylcytosine modification in a large number of targeted regions at single-base resolution. A major advantage of BSPP is that it allows the flexible capture of an arbitrary subset of genomic regions (hundreds to hundreds of thousands of genomic loci) in single-tube reactions. Large number of samples can be processed efficiently and converted into multiplexed sequencing libraries with only three enzymatic steps, without the conventional library preparation procedures. BSPP are applicable to clinical studies, screening cell lines, and for quantifying low abundance regions using deep sequencing.

Key words DNA methylation quantification, Multiplexing, Padlock, Capture, Epigenetics

1 Introduction

A bisulfite padlock probes (BSPP) consist of a common linker sequence connecting two variable capture arms that can anneal to two adjacent genomic regions of bisulfite treated DNA with a gap size up to hundreds of bases [1, 2] (Fig. 1). The capture arms anneal to genomic regions that have been chemically modified by bisulfite such that all unmethylated cytosines are converted to uracils. The first capture arm from the 3′ end anneals in the forward direction, and the linker sequence, which is noncomplementary, provides space to allow the second capture arm to anneal in the reverse direction to a region upstream of the first capture arm. The gap in between the two arms are "captured" by extension with a thermostable polymerase lacking strand displacement and exonuclease activity from the 5′ end of the second capture arm to the 3′ end of the first. After extension, a ligase that is stable and active at high temperatures joins the extension end to the 3′ end of the probe, creating a circular DNA product. The circular DNA product is then amplified by polymerase chain reaction (PCR) and using a

Jörg Tost (ed.), *DNA Methylation Protocols*, Methods in Molecular Biology, vol. 1708,
https://doi.org/10.1007/978-1-4939-7481-8_19,

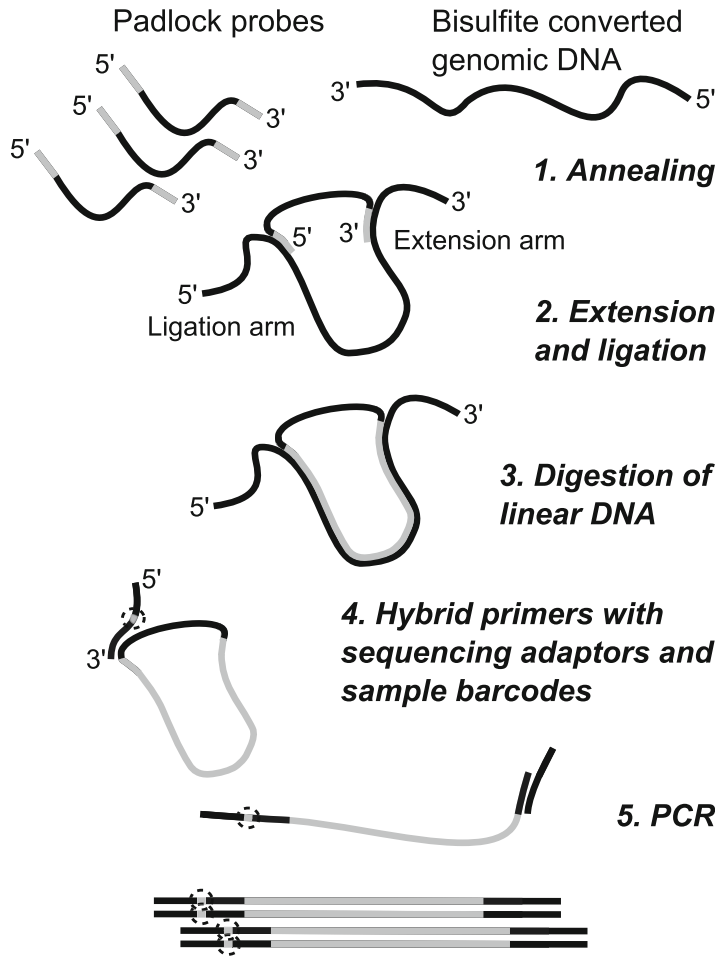

Fig. 1 Bisulfite padlock probe capture. (*1*) The two annealing arms of padlock probes "capture" bisulfite converted DNA during annealing incubation. (*2*) A mixture of dNTPs, polymerase, ligase, and buffer is added to generate a circular product in extension and ligation. (*3*) The genomic DNA and leftover padlock probes are removed by exonucleases. (*4*) Next, hybrid primers are used to label individual samples by appending barcodes (barcode regions are highlighted by *dashed circles*) and to additionally append common adaptors for sequencing. (*5*) PCR amplification is performed on capture products to generate multiple copies of the captured DNA

pair of primers that anneal to the common linker sequence (Fig. 1). The PCR primers contain adaptor sequences and sample barcodes that are compatible with next-gen sequencing instruments. High-throughput sequencing can be performed directly on the capture products, without additional library preparation procedures, to yield information on the methylation of the captured DNA. By design, single molecules such as the Watson and Crick strands are always captured independently. As CpG methylation are expected to be symmetrical, the correlation of CpG sites that are captured on both strands is an estimate of the technical variability within a capture experiment.

Variable capture arm sequences are selected computationally by analyzing the reference genome sequence and fitting user-desired probe constraints. The efficiency of padlock probes, which is the number of molecules captured by each probe sequence, can be estimated using a back-propagation neural network [3]. Efficiency data estimated from previous probe sets was used to train the neural network on the features such as target length, target GC content, binding arm melting temperature, binding arm length, local single-stranded folding energy of the target, and the dinucleotides present at the extension site and ligation sites during probe capture. Padlock probes can be synthesized with standard single columns or on microarrays, with single-columns yielding higher quantities of probes but at a higher up-front cost than array-synthesis. Microarray-based oligo synthesis can produce large numbers (thousands to hundreds of thousands) of oligos but in very limited quantities as a pool. To increase the quantity, bisulfite padlock probes can be synthesized with common adaptors in order to produce functional probes by PCR amplification and enzymatic removal of the adaptors.

Off-target annealing is a primary concern in bisulfite padlock probe capture. During capture, a circular product can only form when a pair of capture arms anneals adjacently to a single molecule with the 5′ and 3′ ends facing each other. Off-targeting annealing primarily leads to lower efficiency when capture arms anneal in multiple locations that do not lead to circular products. If capture arms anneal in off-target locations that leads to a circular product, this would lead to sequences that may not be applicable for analysis, however, this effect can be minimized by ensuring that pairs of capture arm sequences do not overlap with repeats. A minimum length of 25 bp for the capture arm is typically required to obtain melting temperatures above 50 °C, but the combined length of both capture arms should not exceed 60 bp as this would lead to more nonspecific annealing.

Steps for analysis of sequencing reads are similar to the steps taken for whole genome bisulfite sequencing (WGBS, *see also* Chapters 5, 7, and 9) or reduced representation bisulfite sequencing (RRBS, *see* Chapter 8) using readily available tools for next-gen sequence analysis [4, 5]. Similarly, genetic alterations such as single nucleotide polymorphisms (SNPs) can be determined from BSPP data. Furthermore, allele-specific methylation analysis can be performed by capturing regions with heterozygous SNPs [6]. Some considerations for processing BSPP reads include: (1) the necessary trimming from the 5′ end of the reads containing the capture arm sequences and (2) the proportion of clonal reads cannot be identified as all reads map to the same start and end positions. Clonal reads can affect the accuracy of methylation quantification when library complexity is low and too much sequencing was performed on such a library. Appending a unique molecular identifier (UMI) to individual padlock probe oligo is a strategy that allows identification of clonal reads by matching UMIs [7].

2 Materials

2.1 Padlock Probes Design

1. Genome reference sequences in FASTA format are required and can be downloaded from the UCSC Genome Browser. These cannot be in the multi-FASTA format with multiple chromosomes per file.
2. A list of targeted regions in BED format is required to design targets.
3. *ppDesigner* can be downloaded from http://genome-tech.ucsd.edu/public/Gen2_BSPP/ppDesigner/ppDesigner.php.
4. A Mac OS or any other modern Unix-based system is required to run *ppDesigner*.
5. *Perl* is required to run *ppDesigner*. It should already be included in all Unix-based system.
6. *Perl* modules, File::Temp and Sort::Array, are required to run *ppDesigner* and can be downloaded from http://www.cpan.org.
7. *BioPerl* toolkit is required to run *ppDesigner* and can be obtained from http://www.bioperl.org.
8. *Optional*: *UNAfold* software version 3.8. Using *UNAFold* will result in a more accurate prediction of probe efficiency. It is not absolutely required, and it will not change the probe sequence.

2.2 Padlock Probes Preparation (Only for Array-Synthesized Oligonucleotides)

1. Custom synthetic oligonucleotides from service providers such as Agilent, LC Sciences, or CustomArray.
2. 2× KAPA SYBR Fast qPCR master mix (Kapa Biosystems) or similar.
3. Primers for padlock probes amplification (asterisks indicates a phosphorothioate bond).
   (a) pAP1V61U: 5′-G*G*G TCATATCGGTCACTGTU-3′.
   (b) AP2V6: 5′-/5Phos/CACGGGTAGTGTGTATCCTG-3′.
4. Deionized H_2O.
5. 96-well PCR plate.
6. 50 mL disposable reservoir.
7. 200 μL multichannel pipette.
8. Optical adhesive film.
9. 100% Ethanol.
10. 75% Ethanol.
11. 3 M Sodium acetate (NaOAc), pH 5.5.
12. GlycoBlue (15 mg/mL, Thermo Fisher Scientific).
13. TE buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 8.0.

14. QIAquick PCR purification kit (Qiagen).
15. Lambda exonuclease (New England Biolabs, NEB).
16. ssDNA/RNA clean and concentrator column (Zymo Research).
17. USER enzyme mix (NEB).
18. *Dpn*II (50,000 U/mL) (NEB).
19. RE-*Dpn*II-V6 oligo guide: 5′-GTGTATCCTGATC-3′.
20. Real-time thermal cycler.
21. Benchtop centrifuge.
22. 1.5 mL DNA LoBind tube.
23. 0.5 mL microcentrifuge tube.
24. Needle 22 G.
25. 5 mL tube.
26. Nanosep MF 0.2 μm column (PALL).
27. 12-well 6% TBE gel (Thermo Fisher Scientific).
28. 12-well 6% TBE-Urea gel (Thermo Fisher Scientific).
29. 2D-well 6% TBE-Urea gel (Thermo Fisher Scientific).
30. Disposable scalpel #10.
31. 2× TBE-Urea sample buffer (Thermo Fisher Scientific).
32. 10 bp DNA ladder.
33. 10,000× SYBR Gold nucleic acid gel stain (Thermo Fisher Scientific).
34. Benchtop orbital shaker.
35. UV transilluminator.
36. NanoDrop Spectrophotometer (NanoDrop Technologies Inc.) or similar.
37. Qubit fluorometer (Thermo Fisher Scientific).
38. XCell SureLock electrophoresis system (Thermo Fisher Scientific).
39. Electrophoresis power supply.
40. Gel Doc Imager (Bio-Rad) or similar.

2.3 Sample Preparation and Bisulfite Treatment

1. DNeasy blood and tissue kit (Qiagen).
2. TE buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 8.0.
3. Deionized H2O.
4. EZ-96 DNA Methylation-Lightning MagPrep (Zymo Research).
5. 1.5 mL microcentrifuge tube.
6. Thermomixer.

7. Benchtop centrifuge.
8. 0.2 mL PCR tube strips.
9. Qubit dsDNA HS assay kit (Thermo Fisher Scientific).
10. Qubit ssDNA assay kit (Thermo Fisher Scientific).
11. Qubit assay tube (Thermo Fisher Scientific).
12. Qubit fluorometer (Thermo Fisher Scientific).
13. NanoDrop Spectrophotometer (NanoDrop Technologies Inc.) or similar.

2.4 Capture Setup

1. Deionized H_2O.
2. 10× Ampligase Buffer (Epicentre).
3. Functional padlock probes from Subheading 2.2.
4. Bisulfite treated genomic DNA from Subheading 2.3.
5. 96-well PCR plate (Eppendorf).
6. 10 and 20 μL multichannel pipette.
7. Mineral oil.
8. 0.2 mL PCR tube strips.
9. Aluminum seal.
10. Thermocycler.

2.5 Adding KlenTaq, Ligase, and Nucleotides (KLN Mix)

1. Deionized H_2O.
2. 1 mM dNTP mix (NEB).
3. 10× Ampligase Buffer (Epicentre).
4. Hemo KlenTaq (NEB).
5. Ampligase (Epicentre).
6. 0.2 mL PCR tube strips.
7. 10 μL multichannel pipette.

2.6 Exonuclease Digestion

1. Exonuclease I (20 units/μL, Epicentre).
2. Exonuclease III (200 units/μL, Epicentre).
3. 0.2 mL PCR tube strips.
4. 10 μL multichannel pipette.

2.7 Polymerase Chain Reaction

1. Primers for amplification.
   (a) AmpF6.4Sol: 5′-AATGATACGGCGACCACCGAGATC-TACACCACTCTCAGATGTTATCGAGGTCCGAC-3′.
   (b) AmpR6.3 Indexing primers: 5′-CAAGCAGAAGACGG CATACGAGATXXXXXXGCTAGGAACGATGAGCCT-CCAAC-3′ (XXXXXX is the sample barcode, *see* [2] for details).

2. 2× KAPA SYBR Fast qPCR master mix (Kapa Biosystems).
3. 6× Gel loading dye.
4. 50 mL disposable reservoir.
5. 200 μL multichannel pipette.
6. 10 μL multichannel pipette.
7. 96-well PCR plate.
8. Optical adhesive film.
9. PCR cooler.
10. Real-time thermal cycler.
11. Agencourt AMPure XP beads (Beckman Coulter).
12. SPRI plate 96-ring (Beckman Coulter).
13. 75% Ethanol.

2.8 Size Exclusion by Polyacrylamide Gel

1. 5-well 6% TBE gel (Thermo Fisher Scientific).
2. Low DNA mass ladder (Thermo Fisher Scientific).
3. TE buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 8.0.
4. Nanosep MF 0.2 μm column (PALL).
5. Thermomixer.
6. 100% Ethanol.
7. 75% Ethanol.
8. 3 M Sodium acetate (NaOAc), pH 5.5.
9. GlycoBlue (15 mg/mL, Thermo Fisher Scientific).
10. Deionized H_2O.
11. Benchtop orbital shaker.

2.9 Sequencing

1. SolSeq6.3.3 primer: 5′-TACACCACTCTCAGATGTTATCG AGGTCCGAC-3′.
2. SolSeqV6.3.2r primer: 5′-GCTAGGAACGATGAGCCTCC AAC-3′.
3. AmpR6.3IndSeq primer: 5′-GTTGGAGGCTCATCGTTCC TAGC-3′.
4. Illumina sequencing system. Tested systems are MiSeq, HiSeq2000, HiSeq2500 RapidRun, and GAIIx.
5. Illumina Cluster kit (including flowcells).
6. Illumina sequencing kit.

2.10 Data Analysis

1. *BisReadMapper* is required for analysis and can be downloaded from https://github.com/dinhdiep/BisReadMapper. Otherwise, refer to specific instructions for other software packages designed for the analysis of either Whole Genome Bisulfite

Sequencing (WGBS) or Reduced Representation Bisulfite Sequencing (RRBS) (*see also* Chapters 5–9).

2. Genome reference sequences in FASTA format are required and can be downloaded from UCSC Genome Browser.
3. A modern Unix-based system is required to run *BisReadMapper*.
4. *Perl* is required to run *BisReadMapper*.
5. *Trim Galore!* is required: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

   Note that Cutadapt is required also for Trim Galore!
6. b*amUtils* is required: http://genome.sph.umich.edu/wiki/BamUtil.
7. One of the four supported aligners: Supported aligners are *bowtie2* (version bowtie2–2.1.0), *bwa* (bwa-0.7.5a), *SOAP2* (soap2.21release), *LAST* (last-458), or *GEM* (GEM-binaries-Linux-x86_64-core_i3–20,130,406-045632). Note that BWA is recommended for general usage. Refer to *BisReadMapper* README for further instructions.
8. *Samtools* is required for *BisReadMapper* and can be downloaded from http://sourceforge.net/projects/samtools/files/samtools/. Must be version 1.18 or higher.
9. *OPTIONAL*: *Samtools* v1.08 to run the variant caller. Alternatively, use *BisSNP* from http://epigenome.usc.edu/publicationdata/bissnp2011/.
10. *OPTIONAL*: *Perl* module Statistics::LSNoHistory to calculate the forward and reverse correlation.
11. *OPTIONAL*: Latest SNP database for filtering low confident variants from UCSC Genome Browser, i.e., snp147, is http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/snp147.txt.gz.

3 Methods

This protocol can be subdivided into three main phases, which are design, experiment, and analysis. In the design phase, the researcher decides on the genomic targets to capture and design padlock probes to capture those targets. Previous works have established the design parameters that work best for bisulfite padlock probes to enable robust and consistent capture results. Padlock probes can then be ordered through a vendor such as IDT DNA, CustomArray, Agilent, LC Sciences and others. Based on the size of the padlock probe set, the researcher is limited to either array-based or single-tube synthesis methods. Single-column oligonucleotide synthesis is more cost-effective for padlock probe sets of sizes

1–1000, while array-based methods are more cost effective for padlock probe libraries of sizes several thousands and above. The cost per padlock probe is higher for single-tube methods; however, the yields and purity are often better and allows bypassing the additional steps required for the preparation of the padlock probes. Padlock probe preparation is required for libraries from array-based synthesis and take up to a week to complete, however, this can be performed in large batches and can be automated. Once padlock probes are ready, it is possible to move to the experiment phase, which involves the processing of samples. Generally, up to 96 or 384 samples may be processed in parallel in 96-well or 384-well plates to minimize batch-to-batch variability. Each sample is prepared for capture in separate tubes. To prepare for the analysis phase, the samples are prepared for sequencing with a two-step PCR protocol that enables multiplexing of individual samples using barcoded primers and pooled size exclusion to enrich for targeted fragment based on the expected fragment sizes. In the analysis phase, the sequencing library is sequenced and then analyzed with next-generation sequencing analysis software.

3.1 Padlock Probes Design

1. Ensure that individual reference sequences (FASTA) files are placed in a common directory.
2. Convert the target list BED file to a target file in the format required. The file should be tab-delimited and have four required columns (1) the unique ID for each target region, (2) the FASTA filename, such as chr22 (for chr22.fa), (3) The starting position, and (4) the final position. The final fifth column can be the required strand to capture. If strand is not indicated, the program will pick the more efficient probes from either strand. An example target file is given in the Example folder.
3. Generate a job file in the format required. An example job file is given in the Example folder. All of the parameters must be given. The *unafold_path* variable can be set to "NA" if *using_unafold=0*. All paths must be full paths (*see* **Notes 1–5** for important considerations on choosing parameter values).
4. Run the ppDesigner.pl script. See the README for *ppDesigner* for specific usage instructions.
5. Run the primer2padlock.pl script. The *maxH1H2len* is a numerical value and should be the same as the *H1_plus_H2_len* variable from the job file. Indicate "*array*" to generate probes sequences that contain amplification adapters for array synthesis of probes. *See* the README for *ppDesigner* for specific usage instructions.
6. The probe sequences are now ready to be synthesized. It is recommended to randomize the order of the sequences so that

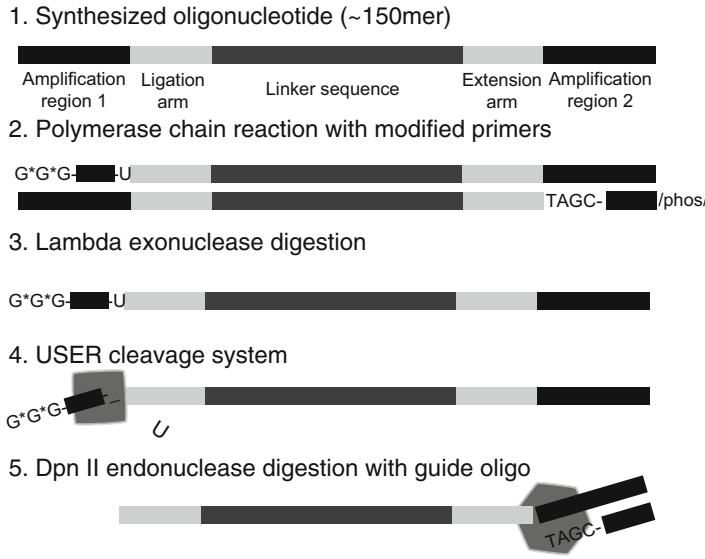

Fig. 2 Bisulfite padlock probe preparation. (*1*) Schematic view of a synthesized oligonucleotide. (*2*) Amplification of oligonucleotides is carried out using a forward primer containing multiple phosphorothioate bonds (marked with a *star*) to prohibit 5′ digestion by lambda exonuclease, and a uracil base at the 3′ end and a reverse primer that is phosphorylated at the 5′ end to promote lambda exonuclease digestion and a *Dpn*II recognition site at the 3′ end. (*3*) Digestion of the reverse complement strand is carried out by lambda exonuclease to generate single stranded probe. (*4*) The 5′ adaptor is cleaved by the USER cleavage system, which contains Uracil DNA glycosylase (UDG, not pictured) that catalyzes the removal of the uracil base and Endonuclease VIII (pictured as a *square*) that breaks the phosphodiester backbone. (5) The 3′ adaptor is cleaved by *Dpn*II (pictured as a *hexagon*) with the help of a guide oligonucleotide to make the cleavage region double-stranded

technical effects such as bad quality spots on the arrays will affect probes in a random manner and do not appear as systematic errors.

3.2 Padlock Probes Preparation

In this section, we describe the preparation of functional padlock probes from libraries of synthetic oligonucleotides (Fig. 2). The libraries of oligonucleotides (typical length of 150 nt and flanked by common amplification adaptors) could be produced by ink-jet printing on programmable microarrays. A first round of PCR, called "expansion PCR," produces the necessary templates for subsequent PCR steps. Amplicons are not functional padlock probes and have to be made functional via a few enzymatic steps that in total can take as little as one full day to accomplish. To purify the functional padlock probe from incompletely digested oligonucleotides, we perform size selection on denaturing 6% TBE-urea PAGE gels.

3.2.1 Expansion PCR

1. Perform the expansion PCR in two reactions (100 μL each) with 1–100 nM template oligonucleotides, 400 nM each of pAP1V61U and AP2V6 primers, and 1× KAPA SYBR Fast

qPCR master mix. The conditions of expansion PCR are as follows: 95 °C for 30 s, 15 cycles of 95 °C for 10 s, 55 °C for 20 s, and 70 °C for 30 s, and final extension at 70 °C for 2 min. The qPCR is monitored and terminated right before fluorescent levels reach plateau (*see* **Note 6**).

2. Purify amplicons with two QIAquick columns following the manufacturer's protocol and elute each column with 50 μL EB buffer (from the QIAquick PCR purification kit).
3. Pool eluted amplicons and measure concentration with NanoDrop.
4. Dilute amplicons to 10 nM and use as template for production PCR.

3.2.2 Production PCR

1. Set up production PCR mix for 100 reactions for 96-well plate (100 μL each) in a 50 mL disposable reservoir. PCR conditions including 0.02 nM of template oligonucleotides, 400 nM each of pAP1V61U and AP2V6, and 1× KAPA SYBR Fast qPCR master mix and amplify at 95 °C for 30 s, 15 cycles of 95 °C for 10 s, 55 °C for 20 s, and 70 °C for 30 s, and final extension at 70 °C for 2 min.
2. Precipitate amplicons in 5 mL tube by pooling 1.2 mL of amplicons to each tube. In each tube, add 120 μL of 3 M NaOAc, pH 5.5 (0.1× volume), 4 μL of GlycoBlue, and 3 mL of 100% ethanol (2.5× volume) and incubate at −80 °C for at least 20 min.
3. Centrifuge at 2700°× *g* for 30 min at 4 °C. Discard supernatant and wash the pellet with 1 mL 75% ice cold ethanol.
4. Dry DNA pellet and resuspend each with 150 μL H_2O.
5. Repurify amplicons with 12 QIAquick columns following the manufacturer's protocol.
6. Elute each column with 50 μL EB buffer and pool the purified amplicons.
7. Measure the concentration with NanoDrop.

3.2.3 Removal of Amplification Adapters

1. Digest the bottom strand of oligonucleotides with lambda exonuclease by incubating 15–20 μg of purified amplicons with 50 units of lambda exonuclease (5 U/μL) in 1× lambda exonuclease buffer in total volume 150 μL at 37 °C for 1 h.
2. Purify the resulting single-strand amplicons with a Zymo ssDNA/RNA clean and concentrator column.
3. Measure the concentration of purified single-strand amplicons with a NanoDrop.
4. To remove the 5′ end amplification adapter, digest 3–5 μg of single-strand amplicons with 5 units of USER (1 U/μL) in 1×

*Dpn*II buffer in total volume 80 μL at 37 °C for 1 h. The USER digested DNA is ~130 nt.

5. To remove the 3′ end amplification adapter, add 5 μL of 100 μM RE-*Dpn*II-V6 guide oligo, 2 μL of 10× *Dpn*II buffer, and 8 μL H_2O to USER digested DNAs, denature at 94 °C for 2 min, decrease temperature to 37 °C for 3 min, add 250 units of *Dpn*II (50 U/μL), incubate at 37 °C for 2 h, and heat inactivate *Dpn*II at 75 °C for 20 min. The USER/*Dpn*II digested DNA size is ~110 nt.
6. Purify USER/*Dpn*II digested DNAs with a Zymo ssDNA/RNA clean and concentrator column.
7. Measure the concentration with a NanoDrop.

3.2.4 Padlock Probe Size Exclusion by PAGE Size Selection

1. Mix 2–3 μg of USER/*Dpn*II digested DNA with 2× TBE-Urea sample buffer in 1:1 ratio in final volume 80–150 μL in a 200 μL PCR tube. Prepare 10 bp DNA ladder in 1× TBE-Urea sample buffer in total volume of 10 μL.
2. Incubate digested DNA and DNA ladder-loading dye mixture at 75 °C for 15 min. At the same time, prerun a 2D-well TBE-Urea gel at 250 V for 15 min.
3. Quick cool digested DNA and ladder on ice, and flush the well with a 1 mL pipette to remove residual urea in the well before loading DNA.
4. Quickly load DNA and DNA ladder into the well.
5. Run at 250 V in 0.5× TBE buffer for 30 min.
6. Remove the gel from the cassette and stain with 1× SYBR Gold in 0.5× TBE buffer for 8 min on an orbital shaker.
7. Place stained gel on a clean Saran wrap on the UV transilluminator with the long wavelength.
8. Locate the position of USER/*Dpn*II digested probe at ~110 bp according to the 10 bp DNA ladder.
9. Cut the gel with scalpel as close as to the band and minimize exposure time to UV light.
10. Move the gel slide to a clean area and chop the gel piece to small pieces with scalpel and transfer to shearing tubes (one gel goes to two tubes).
11. Shearing apparatus preparation for DNA purification from polyacrylamide gel: Place 0.5 mL tube into a 1.5 mL DNA LoBind tube. Punch 0.5 mL tube with a 22 G needle. Treat the tube under UV light for 15 min if possible.
12. Centrifuge the tube at 20,800 × *g* for 2 min at RT and remove 0.5 mL tube.
13. Add 450 μL of 1× TE buffer.

14. Shake vigorously on a shaker for 60 min at 37 °C.
15. Centrifuge the tube at 20,800 × *g* for 2 min at RT.
16. Transfer the clear supernatant to a Nanosep column and centrifuge at 20,800 × *g* for 1 min at RT.
17. Transfer as much as possible the remaining supernatant to a Nanosep column. Centrifuge at 20,800 × *g* for 1 min at RT.
18. Remove the Nanosep column and transfer supernatant to fresh 1.5 mL LoBind tube (volume ~ 400 μL /tube).
19. Precipitate digested DNA with 0.05 μg/μL GlycoBlue, 0.1× volume of 3 M sodium acetate, pH 5.5, and 2.5× volume of 100% ethanol.
20. Mix well and incubate the tube at −80 °C for at least 20 min.
21. Centrifuge at 20,800 × *g* for 30 min at 4 °C.
22. Discard supernatant and wash with 800 μL 75% ice cold ethanol as previously described.
23. Dry the DNA pellet and dissolve with ~10–15 μL H_2O/tube.
24. Pool dissolved padlock probes to the same tube and mix well.
25. Measure DNA concentration with Qubit ssDNA assay.
26. Verify padlock probe preparation on a 12-well 6% TBE-Urea gel using the same procedures as describe above for running a TBE-Urea denaturing gel.

3.3 Sample Preparation and Bisulfite Treatment

1. Extract genomic DNA from cell or tissue samples using a DNeasy blood and tissue kit following the manufacturer's instructions.
2. Measure genomic DNA concentration with Qubit dsDNA HS assay kit and verify genomic DNA quality with NanoDrop.
3. Perform bisulfite conversion on approximately 500 ng to 1 μg of genomic DNA using the EZ-96 DNA Methylation-Lightning Magprep kit in 96-deep-well plate following the manufacturer's protocol and elute bisulfite-treated genomic DNA with 25 μL elution buffer.
4. Measure concentration of bisulfite-converted genomic DNA with Qubit ssDNA assay kit.

3.4 Capture Setup

1. Set up capture reaction in total 20 μL in 96-well plate with the following components: 100–200 ng of bisulfite-treated genomic DNA, normalized amount of BSPP to target based on optimization (*see* **Note 7**), and 1× Ampligase Buffer (*see* **Note 8**).
2. Mix the capture mixture with a 20 μL multichannel pipette and spin down the plate at 1200 × *g* for 1 min.

3. Layer the reaction with 20 μL mineral oil to prevent evaporation during incubation.
4. Seal the plate with aluminum seal and spin down the plate at 1200 × *g* for 1 min.
5. Incubate the capture reactions on thermocycler as the following program: 94 °C for 30 s, gradually lower temperature to 55 °C at 0.02 °C/s in a thermocycler, and let hybridize at 55 °C for 20 h.

3.5 Adding KlenTaq, Ligase, and Nucleotides (KLN mix)

1. Prepare the KLN solution mix as follows: 20% (v/v) Hemo KlenTaq, 0.5 unit/μL Ampligase, 100 μM of dNTP mix, and 1x Ampligase Buffer.
2. Aliquot KLN mix to 8-tube strip and add 2 μL of KLN solution mix to each well with 10 μL multichannel pipette (*see* **Note 9**).
3. Swirl pipette five times to mix the capture reaction without moving the plate from thermal cycler.
4. Continue to incubate at 55 °C for 4–20 h.
5. Heat inactivate enzyme at 94 °C for 2 min.

3.6 Exonuclease Digestion

1. Prepare the exonuclease I/III mix by mixing exonuclease I (20 units/μL) and exonuclease III (200 units/μL) in 1:1 ratio and aliquot equal volumes to a 8-tube strip.
2. Add 2 μL of exonuclease I/III mix to each well with multichannel pipette (*see* **Note 9**).
3. Mix the reaction by swirling pipette around the well five times.
4. Incubate reaction at 37 °C for 2 h.
5. Heat-inactivate exonuclease I/III at 94 °C for 5 min.

3.7 PCR Setup

To enrich for captured DNA, perform PCR to amplify captured DNA in a 100 μL reaction in a 96-well PCR plate.

1. The PCR reaction consists of 10–30 μL of circularized DNA, 200 nM each of AmpF6.4Sol primers and AmpR6.3 indexing primer, and 1× KAPA SYBR Fast qPCR master mix.
2. The PCR program is: 95 °C for 30 s, 8 cycles of 95 °C for 10 s, 58 °C for 30 s, and 72 °C for 20 s, 10–15 cycles of 95 °C for 10 s and 72 °C for 20 s, and a final extension at 72 °C for 3 min. Monitor qPCR during amplification and stop the reaction before the fluorescent curve reaches plateau to prevent over-amplification.
3. Verify the amplified amplicons on a 12-well 6% TBE gel by loading 3 μL of PCR products and run at 250 V in 0.5× TBE buffer for 25 min.

4. Stain the gel with 1× SYBR Gold in 0.5× TBE buffer for 4 min and visualize on a Gel Doc.
5. Purify PCR products with 0.7× volume of AMPure magnetic beads using the standard protocol by the manufacturer and elute with 50 μL TE buffer.
6. Measure the concentration of the purified PCR products with a NanoDrop or BioAnalyzer. BioAnalyzer will give a more accurate reading of the amplicon bands.
7. Pool equimolar ratio of each library into single tube for PAGE size selection.

3.8 Size Exclusion by Polyacrylamide Gel

Perform PAGE size selection of expected DNA size on a 5-well 6% TBE gel. The procedure is similar to that described in Subheading 3.2.4 except that DNA is purified on a TBE gel and no DNA denaturation and prerunning of the gel is performed. Approximately 1 μg can be purified on one gel.

1. Mix pooled sequencing libraries in 1× DNA loading dye and load ~250 ng per well of a 5-well 6% TBE gel. Load Low DNA mass ladder in the middle lane for size estimation.
2. Run at 250 V in 0.5× TBE buffer for 25 min.
3. Stain the gel with 1× SYBR Gold in 0.5× TBE buffer for 4 min as describe in Subheading 3.2.4 in a clean container to avoid library contamination.
4. Place stained gel on a clean Saran wrap on the UV transilluminator with the long wavelength.
5. Locate the position of sequencing libraries by comparing to Low DNA mass ladder.
6. Cut the gel with scalpel as close as to the band of sequencing libraries.
7. Follow the procedures for purifying DNA from polyacrylamide gel as described in Subheading 3.2.4.
8. Quantify sequencing libraries with a Qubit dsDNA HS assay.

3.9 Sequencing

1. Sequencing should be performed by a trained technician either within a Core facility or in-house. The adaptors are designed to be compatible with the Illumina sequencing systems. For other sequencing platforms, the PCR primers need to be modified to be compatible.
2. Paired-end sequencing must be performed for target sizes greater than the limit read length of the sequencer to be able to sequence across the entire target inserts.
3. Use the sequencing primer SolSeq6.3.3 for single-end sequencing or the first read of PE sequencing.

4. Use the sequencing primer AmpR6.3IndSeq for the indexing read.
5. Use the sequencing primer SolSeqV6.3.2r for the second read of PE sequencing.

3.10 Data Analysis

1. Use genomePrep.pl to generate the in silico bisulfite converted references, Cytosines (C) converted to Thymines (T) for Watson, and Guanines (G) converted to Adenines (A) for the Crick strand. Note that bisulfite conversion makes the two strands noncomplementary.
2. Use aligner software to create *index* files from the reference sequence for alignment. Both strands (genome.bis.CT and genome.bis.GA) can be concatenated into one file and only one index needs to be created so long as the aligner and computer RAM can support larger index files.
3. Use samtools "faidx" command to generate the fasta index file from the reference sequence.
4. Generate list_fastq_file or a tab separated table of the files to be processed. Each column in this file represents:

   <sample id> <dir> <read1.fq | read1.fq,read2.fq | sampleID.sam> <phred> <clonal method> <adaptor r1 sequence> <adaptor r2 sequence><library type>. Clonal method could be "none" or "samtools". The supported library types are "RRBS," "WGBS," and "BSPP." Trimming must be done specific to each library type for accurate methylation quantitation. For BSPP libraries, trimming from 5′ end of reads is required to remove capture arm sequences (~27 bases).
5. Generate the list_paths file. An example of a list_paths file is in the Example directory.
6. Run MasterBisReadMapper.pl with list_fastq_files and list_paths as inputs.

4 Notes

1. The maximum target length must take into consideration the desired sequencing platform for the assay. For most applications, a maximum target length of 100 bp should be sufficient. Longer target lengths will require longer sequencing reads or paired-end sequencing to cover the entire targeted region. There is also an inverse correlation between the capture efficiency and the target length.
2. The minimum and maximum target length must be close to enable selection of specific capture products prior to sequencing. Larger differences will lead to more potential nonspecific products being sequenced.

3. The maximum number of CpGs in the capture arm should be limited to 0 or 1. Including more CpGs means that more alternative probes must be synthesized to avoid capture bias towards methylated or unmethylated targets. The capture efficiency of probes can also vary between methylated or unmethylated probes. The presence of CH methylation is negligible in most cell types so we safely assume that they are unmethylated.
4. Minimum H1 and H2 lengths lower than 25 bases should be avoided as this may lead to capture sequences with very low melting temperatures that may not anneal efficiently. H1 plus H2 length should not be longer than 60 bases as this will lead to higher synthesis cost and potentially more nonspecific products.
5. Unique molecular identifiers (UMI) can be attached to padlock probes to allow for single molecules counting and removal of clonal reads [7].
6. There are variations of oligonucleotide quality from different vendors. The concentration of oligonucleotides in expansion PCR should be optimized in small volume (25 μL) by varying concentration between 1, 20, 50, and 100 nM.
7. Probe to target ratio could be varied. Optimization is required to get optimal amount of probes for successful BSPP capture. If the library size is less than ~12,000 probes, the ratio of 1000:1 up to 2000:1 are recommended. If the library size is larger than 12,000 probes, the ratio could be reduced to 500:1 or 200:1.
8. If padlock capture is performed with a small amount of bisulfite converted DNA and a small padlock library size, reducing capture volume to 10 μL can increase capture efficiency.
9. KLN mix and exonuclease I/III mix should be added directly to the capture reaction and not to the layer of mineral oil. This could be done by inserting the pipette tip underneath the mineral oil layer for few seconds before releasing solution mix.

Acknowledgments

We gratefully acknowledge Athurva Gore and Jie Deng for early works on developing the software for padlock probe design and experiments on capture conditions respectively. This work is funded by NIH grants R01GM097253 and R01AG042187. D.D. is supported by a UCSD-CIRM predoctoral fellowship.

References

1. Deng J, Shoemaker R, Xie B et al (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol 27:353–360
2. Diep D, Plongthongkum N, Gore A et al (2012) Library-free methylation sequencing with bisulfite padlock probes. Nat Methods 9:270–272

3. Gore A, Li Z, Fung HL et al (2011) Somatic coding mutations in human induced pluripotent stem cells. Nature 471:63–67
4. Li H, Durbin R (2010) Fast and accurate long-read alignment with burrows-wheeler transform. Bioinformatics 26:589–595
5. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
6. Shoemaker R, Deng J, Wang W et al (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res 20:883–889
7. Hiatt JB, Pritchard CC, Salipante SJ et al (2013) Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res 23:843–854