Journal of Visualized Experiments

Enhanced Reduced Representation Bisulfite Sequencing for assessment of DNA methylation at base pair resolution

--Manuscript Draft--

Manuscript Number:	JoVE52246R4
Full Title:	Enhanced Reduced Representation Bisulfite Sequencing for assessment of DNA methylation at base pair resolution
Article Type:	Invited Methods Article - JoVE Produced Video
Keywords:	epigenetics; bisulfite sequencing; reduced representation; DNA methylation; genomic DNA; 5-methylcytosine
Manuscript Classifications:	7.2.111.87.218: DNA Methylation; 8.1.158.273.180.350.74: Epigenomics; 8.1.158.273.343.350.42: Epigenomics
Corresponding Author:	Francine Garrett-Bakelman Weill Cornell Medical College New York, NY UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author E-Mail:	frg9015@med.cornell.edu
Corresponding Author's Institution:	Weill Cornell Medical College
Corresponding Author's Secondary Institution:	
First Author:	Francine Garrett-Bakelman
First Author Secondary Information:	
Other Authors:	Caroline K. Sheridan
	Thadeous J. Kacmarczyk
	Jennifer Ishii
	Doron Betel
	Alicia Alonso
	Christopher E. Mason
	Maria E. Figueroa
	Ari M. Melnick
Order of Authors Secondary Information:	
Abstract:	DNA methylation pattern mapping is heavily studied in normal and diseased tissues. A variety of methods have been established to interrogate the cytosine methylation patterns in cells. Reduced representation of whole genome bisulfite sequencing was developed to detect quantitative base pair resolution cytosine methylation patterns at GC-rich genomic loci. This is accomplished by combining the use of a restriction enzyme followed by bisulfite conversion. Enhanced Reduced Representation Bisulfite Sequencing (ERRBS) increases the biologically relevant genomic loci covered and has been to profile cytosine methylation in DNA from human, mouse and other organisms. ERRBS initiates with restriction enzyme digestion of DNA to generate low molecular weight fragments for use in library preparation. These fragments are subjected to standard library construction for next generation sequencing. Bisulfite conversion of unmethylated cytosines prior to the final amplification step allows for quantitative base resolution of cytosine methylation levels in covered genomic loci. The protocol can be

	completed within four days. Despite low complexity in the first three bases sequenced, ERRBS libraries yield high quality data when using a designated control lane. Mapping and bioinformatics analysis is then performed and yields data that can be easily integrated with a variety of genome-wide platforms. ERRBS can utilize small input material quantities making it feasible to process human clinical samples and applicable in a range of research applications. The video produced demonstrates critical steps of the ERRBS protocol.
Author Comments:	Dear Editorial staff, Thank you for the opportunity to edit the submitted manuscript and to respond to reviewers' comments. We have uploaded a revised manuscript, updated figures and tables, supplemental code file and response to reviewers. Please do not hesitate to let me know if any further information can be supplied at this time.
Additional Information:	
Question	**Response**
If this article needs to be "in-press" by a certain date to satisfy grant requirements, please indicate the date below and explain in your cover letter.	
If this article needs to be filmed by a certain date to due to author/equipment/lab availability, please indicate the date below and explain in your cover letter.	

Francine E. Garrett-Bakelman
Weill Cornell Medical College
Department of Medicine
Office: 212-746-2932

August 20, 2014

Dear Dr. Nguyen,

We are happy to submit our revised manuscript (JoVE52246R2) 'Enhanced Reduced Representation Bisulfite Sequencing for assessment of DNA methylation at base pair resolution', for your consideration. This paper serves to offer readers a detailed protocol for the preparation of next generation sequencing libraries to interrogate base pair resolution DNA cytosine methylation data at GC-rich genomic loci, inclusive of code for computational analysis downstream.

We have included a full response to the first set of reviews for our paper. Dr. Doron Betel has joined the authorship list, offering critical input (editorial comments and revisions) to the revised submission. The paper has been enhanced by the addition of information on multiplexing approaches for the protocol, manual size selection procedure, and a figure to address several reviewers' comments about data yields. Our manuscript is written per our current protocol. While many protocol enhancements recommended by the reviewers are reasonable, we do not routinely utilize them and have included suggestions in the manuscript text for the readers' consideration of such protocol developments.

We thank Dr. Nandita Singh for the invitation to submit to your journal and for assisting us in the submission process. We welcome any further comments or questions about this manuscript from the reviewers or the editors and we look forward to hearing from you.

Thank you for your time and consideration.

Sincerely,

Francine Garrett-Bakelman, MD, PhD
Instructor in Medicine
Division of Hematology and Medical Oncology
Department of Medicine
Weill Cornell Medical College

Title:
Enhanced Reduced Representation Bisulfite Sequencing for assessment of DNA methylation at base pair resolution

Authors:
Francine E. Garrett-Bakelman*
Department of Medicine
Weill Cornell Medical College
New York, NY
frg9015@med.cornell.edu

Caroline K. Sheridan*
Department of Medicine
Weill Cornell Medical College
New York, NY
cas2038@med.cornell.edu

Thadeous J. Kacmarczyk
Department of Medicine
Weill Cornell Medical College
New York, NY
thk2008@med.cornell.edu

Jennifer Ishii
Department of Medicine
Weill Cornell Medical College
New York, NY
jei2032@med.cornell.edu

Doron Betel
Department of Medicine and Institute for Computational Biomedicine
Weill Cornell Medical College
New York, NY
dob2014@med.cornell.edu

Alicia Alonso
Department of Medicine
Weill Cornell Medical College
New York, NY
ala2035@med.cornell.edu

Christopher E. Mason
Department of Physiology and Biophysics
Weill Cornell Medical College
New York, NY
chm2042@med.cornell.edu

Maria E. Figueroa
Department of Pathology
Michigan State University
Ann Arbor, MI
marfigue@med.umich.edu

Ari M. Melnick
Department of Medicine
Weill Cornell Medical College
New York, NY
amm2014@med.cornell.edu

* These authors contributed equally

Corresponding Author:
Francine E. Garrett-Bakelman
frg9015@med.cornell.edu
Phone number: 212-746-2932

Keywords:
Epigenetics; bisulfite sequencing; DNA methylation; genomic DNA; 5-methylcytosine; high-throughput

Short Abstract:
Enhanced Reduced Representation Bisulfite Sequencing is a method for the preparation of sequencing libraries for DNA methylation analysis based on restriction enzyme digestion combined with cytosine bisulfite conversion. This protocol requires 50 ng of starting material and yields base pair resolution data at GC-rich genomic regions.

Long Abstract:
DNA methylation pattern mapping is heavily studied in normal and diseased tissues. A variety of methods have been established to interrogate the cytosine methylation patterns in cells. Reduced representation of whole genome bisulfite sequencing was developed to detect quantitative base pair resolution cytosine methylation patterns at GC-rich genomic loci. This is accomplished by combining the use of a restriction enzyme followed by bisulfite conversion. Enhanced Reduced Representation Bisulfite Sequencing (ERRBS) increases the biologically relevant genomic loci covered and has been to profile cytosine methylation in DNA from human, mouse and other organisms. ERRBS initiates with restriction enzyme digestion of DNA to generate low molecular weight fragments for use in library preparation. These fragments are subjected to standard library construction for next generation sequencing. Bisulfite conversion of unmethylated cytosines prior to the final amplification step allows for quantitative base resolution of cytosine methylation levels in covered genomic loci. The protocol can be completed within four days. Despite low complexity in the first three bases sequenced, ERRBS libraries yield high quality data when using a designated control lane. Mapping

and bioinformatics analysis is then performed and yields data that can be easily integrated with a variety of genome-wide platforms. ERRBS can utilize small input material quantities making it feasible to process human clinical samples and applicable in a range of research applications. The video produced demonstrates critical steps of the ERRBS protocol.

Introduction:

DNA methylation at cytosine (5-methylcytosine) is an epigenetic mark critical in mammalian cells for a variety of biological processes, including but not limited to imprinting, X chromosome inactivation, development, and regulation of gene expression[1-8]. The study of DNA methylation patterns in malignant and other disorders has determined disease specific patterns, contributions to pathogenesis and biomarker potential[9-17]. There are many protocols that interrogate the epigenome for DNA methylation status. These can be divided into affinity-based, restriction enzyme-based, and bisulfite conversion-based assays that utilize microarray or sequencing platforms downstream. Furthermore, there are a few protocols that bridge these general categories including, but not limited to, Combined Bisulfite Restriction Analysis[18] and Reduced Representation Bisulfite Sequencing (RRBS[19]).

RRBS was originally described by Meissner *et al.*[19,20]. The protocol introduced a step to enrich GC-rich genomic regions followed by bisulfite sequencing, which resulted in quantitative base-pair resolution data that is cost effective[21,22]. The GC-rich regions are targeted by the MspI (C^CGG) restriction enzyme, and cytosine methylation is resolved by bisulfite conversion of cytosines (deamination of unmodified cytosines to uracil), followed by polymerase chain reaction (PCR) amplification. RRBS covered the majority of gene promoters and CpG islands in a fraction of the sequencing required for a whole genome; however RRBS had limited coverage of CpG shores and other intergenic regions of biological relevance. Several groups have published updated RRBS protocols since the original report that improve upon the methodology and resultant coverage of these genomic regions[23-25]. Enhanced Reduced Representation Bisulfite Sequencing (ERRBS) includes library preparation modifications and an alternate data alignment approach[26] when compared to RRBS. ERRBS resulted in a higher number of CpGs represented in the data generated and increased coverage of all genomic regions interrogated[26]. This method has been used to resolve DNA methylation patterns in human patient and other animal specimens[26-30].

The ERRBS protocol described offers details on all steps needed for completion and data was generated using representative human DNA (samples were obtained from previously reported, de-identified patient samples[31], and a CD34+ bone marrow sample from a normal human donor). The protocol includes an automated size selection process, which reduces the processing time per sample and allows for increased accuracy in library size selection. The protocol combines a series of established molecular biology techniques. High molecular weight DNA is digested with a methylation-insensitive restriction enzyme (MspI) followed by end-repair, A-tailing, and ligation of methylated adapters. Size selection of the GC-rich fragments is followed by bisulfite conversion and PCR amplification prior to sequencing. Bisulfite conversion has

been previously described[32] and detailed review of data analysis and applications is beyond the scope of this paper, however recommendations and references are included for the readers' use. The protocol can be performed over four days and is amenable to small input (50 ng or less) material amounts. The protocol as described yields data with high coverage per CpG site sufficient not only for differential methylation site and region determinations but also for epigenetic polymorphism detection as described by Landan, et al.[33].

Protocol:

Institutional review board approval was obtained at Weill Cornell Medical College (protocol number 0805009783) and this study was performed in accordance with the Helsinki protocol.

Caution: Please consult material safety data sheets for relevant materials before use (indicated throughout the protocol with "CAUTION"). Several of the reagents used are toxic and appropriate safety measures are advised (personal protective equipment and fume hood).

Note: All steps are performed at room temperature unless otherwise indicated throughout the protocol

1) Preparation and digestion of genomic DNA

1.1) Prepare 50 ng of high quality genomic DNA (> 40 kilobases in size for human DNA) as starting material in 50 microliter (µl) of DNase-free water. Quantify the DNA using a fluorescence-based quantitation assay per manufacturer's recommendations.

1.2) Mix 50 ng of DNA with 2 µl of MspI (100,000 units/milliliter), 10 µl of appropriate 10X reaction buffer, and DNase-free water to bring the reaction to a total volume of 100 µl.

1.3) Incubate the reaction at 37 °C in a thermal cycler or water bath for at least 18 hours.

1.4) DNA extraction and precipitation

1.4.1) Purify the digested DNA using a phenol-chloroform extraction.

1.4.1.1) Add 200 µl of 10 millimolar (mM) tris(hydroxymethyl)aminomethane (Tris-Cl) pH 8.0 buffer to each reaction from step 1.3 to bring the volume to 300 µl.

1.4.1.2) Add 150 µl of Tris- Ethylenediaminetetraacetic acid (TE) -saturated phenol and 150 µl of chloroform mix (1:1; CAUTION) to the DNA in a chemical hood and vortex briefly.

1.4.1.3) Centrifuge at 20,800 x g in a microcentrifuge for 10 min at room temperature.

1.4.1.4) Transfer top aqueous phase (approximately 300 µl) from last step into a new tube.

1.4.2) Precipitate the digested DNA using ethanol precipitation.

1.4.2.1) Add 1 µl of glycogen (20 milligrams/milliliter), 1/10 volume (30 µl) of 3 M pH 5.2 sodium acetate and 2.5 volumes (750 µl) of room temperature 100% ethanol. Mix by vortexing at high speed.

1.4.2.2) Centrifuge at 4 °C for 45-60 min at 20,800 x g in a microcentrifuge.

1.4.2.3) Remove ethanol quickly by inverting and dragging each tube over a paper towel with a quick flick of the wrist. Ensure that the DNA pellet remains on the side wall of the tube.

1.4.2.4) Add 600 µl of room temperature 70% ethanol to each tube. Visualize the pellet as it dislodges from the tube side. Mix gently by inverting the tube five times.

1.4.2.5) Centrifuge at 4 °C for 45-60 min at 20,800 x g in a microcentrifuge.

1.4.2.6) Remove the ethanol by quickly inverting and dragging each tube over a paper towel with a quick flick of the wrist.

1.4.2.7) Remove as much ethanol as possible by carefully removing any residual volume. Use a vacuum apparatus with a DNase-free 10 µl pipette tip at the end, and allow the pellet to air dry if necessary to fully eliminate any further residual ethanol.

Note: Avoid over drying the pellet since that will reduce the solubility upon resuspension in the next step.

1.4.2.8) Resuspend the DNA pellet into 30 µl of 10 mM Tris-Cl, pH 8.5 (Buffer EB).

Note: To avoid re-annealing of the “sticky” CG overhangs created by the MspI digestion, it is necessary to complete the protocol through the ligation step on the same day (protocol step 4).

2) End-repair

2.1) Transfer the 30 µl of MspI digested DNA (from step protocol 1.4.2.8) into a PCR tube on ice (4 °C) and add the reagents from Table 1.

2.2) Incubate the end-repair reaction for 30 min at 20 °C in a thermal cycler with the heat lid on.

2.3) Purify the DNA products to remove unincorporated dNTPs and other reaction

reagents using a commercial column-based kit that binds of DNA in high-salt buffer and elutes DNA in low-salt buffer conditions per manufacturer's recommendations. Elute the DNA products in 32 µl of Buffer EB.

3) A-tailing

3.1) Transfer the 32 µl of DNA solution from step 2.3 into a PCR tube on ice (4 °C) and add the reagents listed in Table 2.

3.2) Incubate the A-tailing reaction for 30 min at 37 °C in a thermal cycler with the heat lid on.

3.3) Purify the DNA products to remove unincorporated dATPs and other reaction reagents using a commercial column-based kit that binds of DNA in high-salt buffer and elutes DNA in low-salt buffer conditions per manufacturer's recommendations. Elute in 10 µl of Buffer EB.

4) Adapter ligation

4.1) Transfer the 10 µl of A-tailed DNA into a PCR tube on ice (4 °C)

4.2) Add the ligation reaction reagents to the DNA as detailed in Table 3 and adapters as detailed in Table 4.

4.3) Incubate the ligation reaction overnight at 16 °C in a thermal cycler with the heat lid on.

4.4) Purify the ligation products using a solid-phase reversible immobilization (SPRI) bead isolation protocol per manufacturer's recommendations. For example, with Agencourt AMPure XP beads, use a 1.8X ratio of bead volume to sample volume (90 µl of beads for the 50 µl ligation reaction). Elute into 30 µl of DNase-free water.

4.4.1) Store ligated DNA can be stored at -20 °C before proceeding with size selection.

5) Size Selection

Note: Follow either section 5.1 for an automated size selection protocol or section 5.2 for a manual gel extraction size selection protocol. For samples with 25 ng or more of input DNA, an automated size selection protocol (using an instrument such as the Pippin Prep) can be used. Manual gel extraction is necessary for low DNA input amounts of 5-10 ng.

5.1) Size selection using Pippin Prep for samples with 25 ng or higher of input DNA (Figure 2A) .

5.1.1) Create a new protocol on the Pippin Prep for size selection.

5.1.2) Select “2% DF Marker L” as the Cassette. Click the “Use internal standards” button. Verify that the “Ref lane” numbers match the lane numbers.

5.1.3) Select “Range” as the collection mode for each lane. Enter 135 under BP Start, 410 under BP End, and 240 under BP Pause for each lane.

5.1.4) Save the protocol.

5.1.5) Add 10 µl of marker L to each 30 µl sample from the ligation reaction (step 4.3), bringing the total volume to 40 µl. Follow standard Pippin Prep protocol to prepare the gel cassette and instrument.

5.1.6) Remove 40 µl of electrophoresis buffer from the sample wells, and load a 40 µl sample from step 5.1.5 into each of the 5 wells of the dye-free gel cassette.

Note: Dye-free, agarose gel cassettes are critical to size selecting adapter-ligated DNA fragments as the presence of ethidium bromide can alter the migration properties of the forked-adapter bound fragments.

5.1.7) Select the protocol created in step 5.1.1, and start the run.

5.1.8) For each lane used collect 40 µl from the elution module when the run pauses at the 240 base pair (bp) point (lower library fraction: 135-240 bp). The instrument will pause individually for each lane. Repeat steps 5.1.5-5.1.7 for each lane as it pauses at 240 bp.

5.1.9) Wash the elution module with 40 µl of fresh electrophoresis buffer by pipetting up and down three times. Discard the 40 µl of wash buffer. Repeat the wash step two additional times and remove all residual liquid from the elution module. Washing the elution module will decrease the amount of lower fraction DNA carried over into the higher fraction collection.

5.1.10) Add 40 µl of fresh electrophoresis buffer and re-seal the elution module and resume the run.

5.1.11) Collect 40 µl from each elution module into a new tube when the instrument indicates that the elution is complete. This elution contains the higher library fraction (240-410 bp). If running multiple gel cassettes, store size-selected samples at 4 °C before proceeding with bisulfite conversion.

5.2.) Size selection using manual gel extraction (Figure 2B)

5.2.1.) Prepare a 1.5% agarose gel with 0.2 µg/ml ethidium bromide (CAUTION).

5.2.2.) Load 50 bp and 100 bp ladders prepared with loading buffer in adjacent wells on both sides of the gel.

5.2.3.) Add 2 µl of 1X Orange G loading dye to the cleaned-up ligation products. Load the entire volume of each sample into individual wells, skipping at least one well between samples to avoid cross-contamination.

5.2.4.) Run the gel at 3.5 volts per centimeter until ladder is fully separated and the Orange G dye runs to the bottom of the gel (minimum of one hour)

5.2.5.) Slice off the ladders using a clean razor blade and visualize them on a UV transilluminator (CAUTION). Do not expose the samples to UV light. Mark the 150 bp, 250 bp, and 400 bp bands of the ladders using a razor blade or pipette tips. Marking the ladder bands will allow the excision of samples without exposing them to UV light.

5.2.6) Align the ladder slices with the rest of the gel. Using the marked bands as reference, excise one slice containing the 150-250 bp library fraction (lower) and another slice containing the 250-400 bp library fraction (higher). Place each of the two slices into different tubes.

5.2.7) Repeat the excision for all samples prepared using a clean razor blade for each gel lane used.

5.2.8) Purify the lower and higher library fractions using a gel extraction kit following manufacturer's protocols. Elute each sample extracted into 40 µl of EB. To ensure efficient PCR amplification of both the lower and higher library fractions in step 7, maintain the two fractions as independent samples for the bisulfite conversion step.

6) Bisulfite Conversion

6.1) Set up a bisulfite conversion control per manufacturer's recommendations (e.g., Universal Methylated Human DNA Standard kit). Treat the control in the same manner as the samples throughout the bisulfite conversion protocol.

6.2) Perform bisulfite conversion using a commercial kit protocol per manufacturer's recommendations. Use a commercial kit that has limited template degradation and DNA loss during treatment and cleanup, while resulting in high bisulfite conversion rates of the DNA. If using the EZ DNA Methylation Kit, perform the protocol per manufacturer's recommendations with the following exception: incubate the samples in a thermal cycler using the following protocol: 55 cycles: 95 °C for 30 sec, 50 °C for 15 min. Hold at 4 °C.

6.3) Elute the samples in 40 µl of DNase free water. Proceed with the enrichment PCR (step 7) on the same day that the bisulfite conversion is completed. Bisulfite-treated DNA is AU-rich and single-stranded, which reduces its stability.

6.4) Confirm efficient bisulfite conversion of the control used via Sanger sequencing

with non-methylated cytosines greater than 99% converted.

7) Enrichment PCR

7.1) Prepare a PCR master mix using the reagents in Table 5 per library fraction (optimized for the use of FastStart Taq DNA Polymerase).

7.2) Add the PCR master mix to each 40 µl bisulfite-converted library fraction. Mix by pipetting. Divide the 200 µl reaction into four PCR tubes with 50 µl each.

7.3) Amplify the reactions in a thermal cycler with the following protocol: Set the heat lid to 100 ^{o}C. Initialize with a step of 94 °C for 5 min. Run 18 cycles of denaturing, annealing and extension/elongation steps: 94 °C for 20 seconds followed by 65 °C for 30 seconds followed by 72 °C for 1 min. Run a final extension/elongation step of 72 °C for 3 min and hold at 4 °C.

Note: 18 cycles of PCR is recommended for low input material quantities (less than 10 ng) and for first time users of the protocol. The protocol can be adjusted for a lower number of PCR cycles (as low as 14 cycles for 50 ng of input DNA; see Table 6).

8) Purify PCR reactions

8.1) Combine each 4-reaction set (four 50 µl PCR reactions per library fraction).

8.2) Purify the PCR products using a SPRI bead approach (or other approach which can remove unincorporated primers and other PCR reaction reagents) per manufacturer's recommendations. The follow steps have been optimized for Agencourt AMPure XP.

8.2.1) Add 1.7X volume of SPRI beads (340 µl for 200 µl of PCR product) to the lower fraction amplification reactions.

8.2.2) Add 1.1X volume of SPRI beads (220 µl for 200 µl of PCR product) to the higher fraction amplification product.

8.2.3) Purify PCR products per manufacturer's recommendations using 800 µl of 70% ethanol for wash steps.

8.2.4) Elute into 50 µl of DNase-free water and add 1M Tris buffer to bring each library to 10 mM Tris-Cl, pH 8.5. Store libraries at -20 °C.

9) Library Quality Control

9.1) Quantify libraries using a fluorescence-based quantitation assay selective for double-stranded DNA per manufacturer's recommendations. Spectrophotometry-based measurements are not reliable. Expected concentrations measure 10 - 50 ng/µl for the

lower library fraction and 3 - 15 ng/µl for the higher library fraction.

9.2) Assess library sizes and quality using a bioanalyzer instrument and the High Sensitivity DNA Kit. Visualize library products and determine the average size of each library fraction. The lower library fraction typically has an average size between 180 and 210 bp. The higher fraction typically has an average size between 280 and 310 bp.

10) Prepare libraries for sequencing

10.1) Calculate the molarity of each library fraction as follows:

10.1.1) Size each library fraction using the trace obtained in step 9.2. For example, if using a bioanalyzer, use the “region” feature, and cover the beginning and end of the library fraction assessed.

10.1.2) Record the average size of each library fraction in bp.

10.1.3) Calculate molarity (nanomolar; nM) of library fraction using the following formula: nM = $[(ng/1000)/bp*660]*10^9$ where ng is the concentration expressed in ng/µl (as measured in step 9.1) and bp is the average size of the library fraction. For example: with a library that is 326 bp in size, at a concentration of 14.2 ng/µl, the molarity is 66 nM.

10.2) Prepare 10 µl of a 2 nM solution of each library fraction. Dilute the library with DNase-free water.

10.3) Pool the library fractions by combining 10 µl of the 2 nM lower fraction with 10 µl of the 2 nM higher fraction for each sample library prepared. This pool is the final ERRBS library for sequencing.

10.4) Sequence the ERRBS library

10.4.1) Load the libraries with a goal of optimal densities of 600,000 – 650,000 clusters per millimeter squared (suggested loading concentration of 7-8 picomolar).

10.4.2) Sequence the libraries using a minimum of single-read 51 cycles sequencing on a HiSeq 2500 sequencer in high output mode.

10.4.2) Use a designated control lane (see discussion for rationale).

11) Data Analysis

Note: Please refer to the Supplemental code files 1 and 2 for full details of commands and scripts recommended for use.

11.1) Convert the base call files (.bcl files) to individual FASTQ files for each sample

using the software provided by the sequencer's manufacturer (example: CASAVA v1.8.2 for Illumina). If a sequencing core facility is used, this step may be provided and/or performed as an automated step on-board the sequencing computer.

11.2) Filter the sequencing reads for reads that pass quality filtering (see Supplemental code file 1 for details). The CASAVA FASTQ file contains both reads that pass quality filtering and reads that do not pass quality filtering. Use a custom script that utilizes the <is filtered> element of the sequence identifier and keeps reads that have passed quality filtering.

11.3) Trim adapter sequences from 3' end of the sequence reads in the filtered FASTQ files using a software tool that can remove adapter sequences, such as Flexbar [37]. The minimum overlap length (-ao) set to 6, the minimum read length to keep after adapter removal (--m) set to 21, the cutoff for the number of allowed mismatches (-at) set to 2, and defaults for all other parameters. As an alternate to Flexbar, any software which can remove adapters can be used in step 11.3 (examples: cutadapt[38], trimmomatic[39]).

11.4) Use Bismark[41] to align the filtered, adapter trimmed sequence reads to the bisulfite converted reference human genome hg19 (whole genome alignment approach) and determine the methylation context for each cytosine. Bismark is a customized short read mapping tool that aligns bisulfite treated reads to a bisulfite converted genome (where all 'C's are converted to 'T's) and returns methylation calls for cytosines in CpG, CHG, and CHH context. Typically set the seed length (-l) to the read length for alignment accuracy and use defaults for all other parameters.

11.5) Sort aligned reads first by chromosome, then start position, and finally strand, after alignment is complete.

11.6) Use custom scripts (see Supplemental code file 2 for script commands) to iterate over the sorted methylation calls output by Bismark to compute the percent methylation scores of bisulfite converted Cytosines (T's; representing unmethylated C's) and non-converted C's (representing methylated C's) for each cytosine methylation context (CpG, CHG, CHH), retaining only the cytosines that have at least phred quality score of 20 and have at least 10x coverage.

Note: The outputs are a methylation score file for each cytosine context with columns corresponding to: the position, strand, coverage, percent cytosine, and percent thymine, and for CHG and CHH context, a column for the next base (the H).

11.7) Use a custom script (see Supplemental code file 2 for script commands) to compute the conversion rates, from the methylation scores, and output the total of other C's considered (CHG and CHH context), average conversion rate, and median conversion rate. Compute conversion rates for both strands independently as well as summarized over the entire library. Mean conversion rate is the fraction of C's (converted and unconverted) in non-CpG context of the total number of C's. Cytosines in CHG or CHH context are typically unmethylated and therefore present as thymine in

the sequencing data.

11.8) Generate a BAM file of the aligned reads using Bismark's bismark2SAM_v5_xm.pl[41] and SAMTOOLS[42], and generate a wiggle format file using a custom script to convert the CpG methylation calls. The post-processed output can be converted to other formats, such as a bedgraph, and viewed in a genome browser such as the UCSC Genome Browser[43] or IGV[44].

Representative Results:
Figure 1 provides an overview of ERRBS, highlighting key steps, which are explained throughout the protocol described. ERRBS libraries were prepared using 50 ng input DNA.

Evaluate the quality of the libraries prepared. Library production routinely yields fraction sizes of 150-250 bp and 250-400 bp (Figure 3A-C). Slight differences in library size distributions between samples are expected. Note that in both lower and higher library fractions there are very intense DNA sizes, indicative of enrichment of a particular sequence. MspI digestion results in the enrichment of a family of repetitive DNA sequences present in the human genome at 190 bp, 250 bp and 310bp in the ERRBS libraries. These three repeats represent a characteristic signature of an ERRBS library[20] (see Figures 3A-C and 3G). Representative libraries were sequenced on a next-generation sequencer using single-end reads. When loading at the recommended library concentration on an Illumina HiSeq 2500 sequencer, cluster densities of 500,000-700,000 per mm^2 are expected. At this clustering density, 81.6 +/- 3.14% (n = 81) of the clusters pass filter (Figure 4A). Due to the low complexity end of the library inserts (MspI recognition site: C^CGG), intensity values and quality scores recorded during sequencing are highly variable in the first three bases (Figure 4 B-C), however, if an independent control lane is included (see discussion), 85% of bases will have quality scores of 30 or greater (Q30 values; Figure 4D).

Data alignment and cytosine methylation determination as described in the protocol yields base-pair resolution data (Table 7). For the human genome, a 51-cycle single-read sequencing run of an ERRBS library in one lane of a HiSeq 2500 in high output mode regularly generates 153,194,882 +/- 12,918,302 total reads that after quality filtering and adapter trimming yields 152,231,183 +/- 13,189,678 reads for input into the analysis pipeline. Average mapping efficiency for an ERRBS library is typically 62.95 +/- 5.92% with representation of 3,183,594 +/- 713,547 CpGs with a minimum coverage per CpG of 10X and an average coverage per CpG of 84.94 +/- 16.29 (n = 100).

The ERRBS protocol is amenable to multiplexing (see Supplemental file 1: Protocol adaptation for multiplexed sequencing). Data from representative sequencing runs is summarized in Figure 5. Data from multiplexed sequencing runs (51-cycle single-read sequencing run; n = 128 for two libraries per lane; n = 11 for three libraries per lane; n = 11 for four libraries per lane) were compared to a full lane sequencing of an ERRBS library (51-cycle single-read sequencing runs; n = 100) as well as downsampling a single lane to simulate 50%, 33% and 25% of reads per lane (2, 3, and 4 sample

multiplexing per lane respectively; n = 3). As the number of reads per sample decreases with the multiplexing factor, the number of CpGs covered at a minimum coverage of 10X and the coverage per CpG decreases as well (Figure 5 and Table 8). Mean conversion rates of non-CpG sites expected are 99.85 +/- 0.04% (n = 400). Conversion rates lower than 99% may indicate less than optimal bisulfite conversion that can result in high rates of false methylation levels.

Data from an ERRBS library prepared from a representative human genomic DNA was analyzed in R 2.15.2[45] using the methylKit package[26] (see Supplemental code file 1 for command details). The data can be visualized in commonly used genome browsers (Figure 6A). The cytosine methylation data is equally derived from both strands (Figure 6B) and ranges the entire spectrum of potential cytosine methylation levels (Figure 6C). Analysis of technical replicates from a representative human DNA sample yields high concordance between the data results (Figure 6D) and covers CpGs in a broad spectrum of genomic loci (Figure 6E and F and as previously described[26]). While technical replicas will yield high R^2 values (greater than 97%), biological replicas will yield R^2 values ranging from 0.92 – 0.96[26], and comparing different human cell types will yield R^2 values lower than 0.86 (data not shown).

Figure 1: Flow chart of the ERRBS protocol steps. Chart represents steps, which can be completed in a traditional work day. * indicates a potential pause point (immediately following ligation clean up and before size selection, protocol step 5) at which samples can be frozen at -20 °C before proceeding with the duration of protocol.

Figure 2: Size selection protocol. A) Screen shot of settings used in the ERRBS Pippin Prep protocol (see protocol section 5.1.2 – 5.1.6): (1) Select Cassette type. (2) Select standard to be used. (3) Select the collection mode for each lane. (4) Enter the collection bp ranges. (5) Save the protocol. B) Stages of the manual gel extraction used in protocol section 5.2: (1) Visualized gel ladders. (2) Marked Sizes for size selection using a razor blade. (3) Image of excised samples (lower fraction: 150 – 250 bp and higher fraction: 250 – 400 bp).

Figure 3: Quality control results for representative ERRBS libraries prepared from human DNA samples using a bioanalzyer machine. A) Gel-like image showing a standard ladder (1), lower library fraction (135 - 240 bp fraction from Pippin Prep); 2) and the higher library fraction (240 - 410 bp fraction from Pippin Prep); 3). B) Bioanalyzer electropherogram of the expected lower library fraction. C) Bioanalyzer electropherogram of the expected higher library fraction. D - F) Representative data from a poor quality library prep. Gel-like image (D) of the standard ladder (1), lower library fraction (2) and the higher library fraction (3). The band at 150 bp marked with an arrow indicates excessive amounts of adapter. Electropherogram of the lower (E) and higher library fractions (F) with the excess adapter peaks at 150 bp (marked with arrows). G) Bioanalyzer electropherogram of a pooled ERRBS library for sequencing. Red trace represents a high quality pooled library with equal representation of higher and lower fractions. Blue trace represents a pooled library not adequate for sequencing due to a lack of equal representation of the higher and lower fractions.

Figure 4: Sequencing charts for a representative ERRBS 51-cycle single-read sequencing run on a HiSeq 2500 sequencer in high output mode. A) Cluster densities (K/mm^2 = 1,000 clusters per millimeter squared; blue) and cluster densities passing filter (green) in two lanes with ERRBS libraries. B) Typical intensities seen in the first 30 cycles in a lane with an ERRBS library. Note the CGG signature from MspI digestion in the intensities of the first three cycles. C) Percentage of bases with a quality score of 30 or higher (%>Q30) for each cycle in one ERRBS lane. D) Quality score distribution for all cycles in one ERRBS lane. Blue = less than Q30, Green = greater than or equal to Q30. In this lane, 84.7% of bases had quality scores of 30 or higher.

Figure 5: Sequencing output results. Box plots of experimental data from multiplexed and single sample per lane sequencing runs (displayed as green boxes) and of data derived by simulated downsampling from sequencing runs of three ERRBS libraries (displayed as blue boxes; sampled five times for each sequencing run) from 51-cycle single-read sequencing runs. The multiplexing factor corresponds to the number of ERRBS libraries sequenced per lane. 1 = whole lane or 100% of reads and represents data from a single ERRBS library per lane; 2 = 50% of lane and represents data from two ERRBS libraries per lane; 3 = 33% of a lane and represents data from three ERRBS libraries per lane; and, 4 = 25% of a lane and represents data from four ERRBS libraries per lane. A) The read counts, or number of sequences analyzed, per multiplexing factor. B) The number of CpG's covered by the sequencing data per multiplexing factor. C) The mean coverage per CpG per multiplexing factor.

Figure 6: Representative data from an ERRBS library prepared from human genomic DNA. A). University of California, Santa Cruz (UCSC) genome browser[43] image of representative data from an ERRBS sequencing lane. The y-axis scale bar represents 0-100% methylation at each cytosine covered with a minimum of 10X. The top custom track represents the forward strand and the lower custom track represents the reverse strand. Shown is chr12:6,489,523-6,802,422 (hg19) inclusive of refseq genes and CpG islands within this genomic region. B) Distribution histograms of CpG coverage along forward and reverse strands in a representative human CD34+ bone marrow sample. C). Distribution histogram of CpG methylation levels along both strands in a representative human CD34+ bone marrow sample. D). Correlation plot of CpG methylation levels from a representative technical replica of a human DNA sample. E). Pie chart illustrating the proportions of CpGs covered in ERRBS which annotated to CpG islands (light green), CpG shores (gray) and other regions (white) in a representative sample prepared from human genomic DNA. F). Pie chart illustrating the proportions of CpGs covered in ERRBS which annotated to gene promoters (red), exons (green), introns (blue) and intergenic regions (purple).

Table 1: End repair reaction reagents. Reagent names and quantities used in the end repair reaction (protocol step 2.1)

Table 2: A-tailing reaction reagents. Reagent names and quantities used in the A-tailing reaction (protocol step 3.1)

Table 3: Adapter ligation reaction reagents. Reagent names and quantities used in the adapter ligation reaction (protocol step 4.2)

Table 4: Oligos used in the ERRBS protocol. List of oligos used throughout the ERRBS protocol in the ligation reaction (protocol step 4) and PCR amplification steps (protocol step 7).

Table 5: PCR reaction reagents. Reagent names and quantities used in the PCR amplification reaction (protocol step 7.1)

Table 6: Protocol step modifications for input material quantities ranging from 5 – 50 ng. Several steps throughout the protocol require modification of reagent quantities used to generate high quality libraries from various quantities of starting materials. Changes to key reagent quantities are included here. Adjust buffer and water volumes in reactions accordingly.

Table 7: Representative ERRBS data. After data alignment and cytosine methylation determination, base pair data is obtained. For each CpG covered, the alignment protocol as described will determine the genomic coordinate (columns: chr = chromosome, Base and Strand), the coverage rate of the specific locus (Coverage), and the rate of detection cytosine versus thymidine as percent (freqC and freqT respectively).

Table 8: Representative parameters from sequencing single and multiplexed ERRBS libraries. Shown is data per lane from 51-cycle single-read sequencing runs: mean and standard deviations of uniquely aligned reads, number of CpGs covered and coverage per CpG site obtained from sequencing single ERRBS libraries per lane (n = 100), two ERRBS libraries per lane (n = 128), three ERRBS libraries per lane (n = 11), and four ERRBS libraries per lane (n = 11).

Discussion:

The protocol presented yields base-pair resolution data of cytosine methylation at biologically-relevant genomic regions. The protocol as written is optimized for 50 ng of starting material, however, it can be adapted to handle a range of input material (5 ng or more)[26]. This will require adjustments of some of the protocol steps as seen in Table 6. The ERRBS libraries are amenable to paired end sequencing and further genomic coverage can also be accomplished by sequencing reads longer than 51 cycles. Multiplexed sequencing will offer a lower cost protocol per sample, however, this will result in reduced coverage per CpG site represented in the data (Figure 5 and Table 8), and will not yield sufficient depth of coverage to perform analyses which require high coverage per CpG site (e.g. as described by Landan et al.[33]). Finally, this protocol (or any bisulfite-based protocol) cannot distinguish between methyl-cytosine and hydroxymethyl-cytosine[46,47]. However, the data generated can be integrated with other protocol results[48,49] to delineate the different modifications, and other cytosine modifications recently reported[50], should they be of interest.

High quality libraries will appear as shown in Figure 3A-C, and once pooled for sequencing yields a trace as shown in Figure 3G (red trace) representing equal molar contributions from both library fractions. Library preparation failure can result from any step during the procedure. If degraded DNA is processed it will result in libraries that are not enriched in MspI fragments and hence in low CpG coverage using the sequencing parameters described in this protocol. If an enzyme is non-functional or inadvertently excluded from one of the reactions, the protocol will not yield the expected library. If the ligation reaction is inefficient, adaptors are at a higher concentration than expected, and/or the primers concentration used is a limiting reagent for the final amplification steps, library failure can occur. Excess adapters (seen as a peaks at ~150 bp in bioanalyzer results; Figure 3D-F) in the library will also interfere with sequencing due to the indiscriminate clustering of both the library and excess adapters. While such a library may sequence apparently normally, a significant portion of the reads will be merely adapter sequences. If excess adapters are observed in a library, it is best to repeat the library preparation if material is available using optimal input material to adapter quantity ratios. Finally, to ensure efficient PCR amplification of the libraries, the lower and higher library fractions are maintained as separate samples throughout the bisulfite conversion and PCR enrichment steps. Failure to do so yields differential efficiency of amplification during the PCR reaction of higher and lower fractions (as seen in Figure 3G blue trace) and the potential for unequal representation of the respective genomic loci covered in each library fraction during sequencing. The user may opt to include a quantitative PCR step immediately after the bisulfite conversion for further titration of optimal PCR cycles needed to amplify the libraries being generated.

ERRBS library preparation protocol has several key steps in which specific reagents are recommended. At the end-repair step, the use of a four-nucleotide dNTP mix allows for end-repair of any products not containing the CG overhang, such as those resulting from MspI enzymatic star activity and sheared DNA fragments present in the original DNA sample. This results in improved CpG representation in the results. At the ligation step it is critical to use a high concentration ligase (2,000,000 units/milliliter) and methylated adapters to ensure that the ligation reaction is efficient and that the bisulfite conversion does not influence the adapter sequences essential for accurate data alignment. At the PCR step, using a polymerase capable of amplifying bisulfite-treated GC-rich DNA fragments is necessary for high specificity. Finally, to ensure elimination of excess adapters and primers, SPRI bead purification (for example: Agencourt AMPure XP) is recommended rather than column based assays for ligation and PCR product isolations.

In order to generate high quality data, it is important to ensure efficient bisulfite conversion. The control presented offers the user the ability to determine conversion efficiency prior to sequencing. As an alternative, a non-human DNA such as lambda DNA can be used as an internal control (spike-in). Due to the differences in species, this type of a control can be directly included in downstream sequencing (e.g. as used by Yu, et al.[34]). However, if the spike-in is utilized, it cannot be used to determine conversion efficiency prior to library sequencing unless uniquely amplified and

independently sequenced prior to library sequencing. The conversion rates determined are based on the methylation status at non-CpG sites. This may not be appropriate for use in the context of high cytosine methylation in non-CpG context (for example embryonic stem cells) and parallel samples or other means of assessing for conversion efficiency can be utilized for this purpose.

There are a few caveats to address that are unique to the sequencing of ERRBS libraries. The first three bases of the library fractions sequenced are nearly uniformly non-random due to the MspI recognition cut site (C^CGG; see Figure 4B, C). This results in the potential for significant data loss due to low quality reads resulting from poor cluster localization in spite of apparent high cluster density during sequencing. To overcome this barrier, include a high complexity library in an independent lane (PhiX control or other library type) as a dedicated control lane. High complexity libraries have ends containing a balanced representation of A, C, T and G in the first four bases sequenced. Suitable control lanes include libraries such as RNA-seq, ChIP-seq, whole genome sequencing, or a control offered by the sequencing machine manufacturer (e.g. PhiX Control v3). When designated as a control lane for the respective sequencing run, it can serve as the basis for the matrix generation which is utilized during the first four bases of sequencing to detect cluster positions. The higher quality reads captured will raise the mean coverage per CpG site by 5.2 (n = 4). Alternatively, this technical difficulty can also be overcome using a dark sequencing approach as previously described[23]. Other sequencing criteria follow standard operating procedures per manufacturer's protocols. Finally, the coverage per CpG chosen for data analysis will be guided by the user and in part by the biological questions of interest. 10X coverage threshold affords a high coverage analysis approach, however this threshold can be lowered should that be of interest.

A full discussion of ERRBS data analysis is beyond the scope of this article, however, differentially methylated cytosines and regions can be determined using open source tools [31,51-53]. Additional analysis considerations and approaches have been well-described[54,55], and the reader is encouraged to search the literature for tools most appropriate to the analysis planned.

Compared to other published methods, ERRBS offers a four-day protocol which when performed as described yields high rates of reproducibility. It has been validated compared to the gold standard MassARRAY EpiTYPER[26], is cost-effective for high coverage data, and is adaptable for various input material amounts (favorable for clinical sample processing and other cell types of low frequency) and sequencing approaches. It offers base-pair resolution at biologically relevant loci and can be used in integrative analyses with other techniques profiling genome-wide transcription factor binding, chromatin remodeling, epigenetic marks and other cytosine modifications of interest. ERRBS data use in such studies can contribute to a comprehensive molecular approach and allow for high dimensional analyses in the study of biological models and human disease.

Acknowledgements:

We thank all the authors of the original ERRBS report. We thank Mame Fall for technical assistance. We acknowledge the Weill Cornell Medical College Epigenomics Core for technical services and assistance. The work was supported by a Sass Foundation Judah Folkman Fellowship, an NCI K08CA169055 and ASH-AMFDP12005 to FGB, NIH R01HG006798 and R01NS076465, funding from the Irma T. Hirschl and Monique Weill-Caulier Charitable Trusts and STARR Consortium (I7-A765) to CEM, and an LLS SCORE grant (7006-13) to AMM.

Disclosures:
The authors have no conflicts of interest to disclose.

References:
1 Jones, P.A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. **13** (7), 484 - 492, doi: 10.1038/nrg3230 (2012).
2 Barlow, D.P. Genomic imprinting: a mammalian epigenetic discovery model. *Annual Review Of Genetics*. **45**, 379 - 403, doi: 10.1146/annurev-genet-110410-132459 (2011).
3 Thiagarajan, R.D., Morey, R., Laurent, L.C. The epigenome in pluripotency and differentiation. *Epigenomics*. **6** (1), 121 - 137, doi: 10.2217/epi.13.80 (2014).
4 Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*. **447** (7143), 425 - 432, doi: 10.1038/nature05918 (2007).
5 Hartnett, L., Egan, L.J. Inflammation, DNA methylation and colitis-associated cancer. *Carcinogenesis*. **33** (4), 723 - 731, doi: 10.1093/carcin/bgs006 (2012).
6 Smith, Z.D., Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet*. **14** (3), 204 - 220, doi: 10.1038/nrg3354 (2013).
7 Li, E., Bestor, T.H., Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*. **69** (6), 915 - 926 (1992).
8 Okano, M., Bell, D.W., Haber, D.A., Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. **99** (3), 247 - 257 (1999).
9 Feinberg, A.P. Phenotypic plasticity and the epigenetics of human disease. *Nature*. **447** (7143), 433 - 440, doi: 10.1038/nature05919 (2007).
10 Bock, C. Epigenetic biomarker development. *Epigenomics*. **1** (1), 99 - 110, doi: 10.2217/epi.09.6 (2009).
11 Laird, P.W. The power and the promise of DNA methylation markers. *Nat Rev Cancer*. **3** (4), 253 - 266, doi: 10.1038/nrc1045 (2003).
12 How Kit, A., Nielsen, H.M., Tost, J. DNA methylation based biomarkers: practical considerations and applications. *Biochimie*. **94** (11), 2314 - 2337, doi: 10.1016/j.biochi.2012.07.014 (2012).
13 Mikeska, T., Bock, C., Do, H., Dobrovic, A. DNA methylation biomarkers in cancer: progress towards clinical implementation. *Expert Review Of Molecular Diagnostics*. **12** (5), 473 - 487, doi: 10.1586/erm.12.45 (2012).
14 Gyparaki, M.T., Basdra, E.K., Papavassiliou, A.G. DNA methylation biomarkers as diagnostic and prognostic tools in colorectal cancer. *Journal of Molecular Medicine*. **91** (11), 1249 - 1256, doi: 10.1007/s00109-013-1088-z (2013).

15 Figueroa, M.E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell*. **17** (1), 13 - 27, doi: 10.1016/j.ccr.2009.11.020 (2010).
16 Heyn, H., Mendez-Gonzalez, J., Esteller, M. Epigenetic profiling joins personalized cancer medicine. *Expert review of Molecular Diagnostics*. **13** (5), 473 - 479, doi: 10.1586/erm.13.36 (2013).
17 Kulis, M., Esteller, M. DNA methylation and cancer. *Advances in Genetics*. **70**, 27 - 56, doi: 10.1016/B978-0-12-380866-0.60002-2 (2010).
18 Xiong, Z., Laird, P.W. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res*. **25** (12), 2532 - 2534 (1997).
19 Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res*. **33** (18), 5868 - 5877, doi: 10.1093/nar/gki901 (2005).
20 Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc*. **6** (4), 468 - 481, doi: 10.1038/nprot.2010.190 (2011).
21 Bock, C. *et al.* Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol*. **28** (10), 1106 - 1114, doi: 10.1038/nbt.1681 (2010).
22 Harris, R.A. *et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol*. **28** (10), 1097 - 1105, doi: 10.1038/nbt.1682 (2010).
23 Boyle, P. *et al.* Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol*. **13** (10), R92, doi: 10.1186/gb-2012-13-10-r92 (2012).
24 Chatterjee, A., Rodger, E.J., Stockwell, P.A., Weeks, R.J., Morison, I.M. Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. *Journal of Biomedicine & Biotechnology*. **2012**, 741542, doi: 10.1155/2012/741542 (2012).
25 Lee, Y.K. *et al.* Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples. *Biological Procedures Online*. **16** (1), 1, doi: 10.1186/1480-9222-16-1 (2014).
26 Akalin, A. *et al.* Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet*. **8** (6), e1002781, doi: 10.1371/journal.pgen.1002781 (2012).
27 Hatzi, K. *et al.* A Hybrid Mechanism of Action for BCL6 in B Cells Defined by Formation of Functionally Distinct Complexes at Enhancers and Promoters. *Cell Reports*. **4** (3), 578 - 588, doi: 10.1016/j.celrep.2013.06.016 (2013).
28 Will, B. *et al.* Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. *Nature Immunology*. **14** (5), 437 - 445, doi: 10.1038/ni.2572 (2013).
29 Lu, C. *et al.* Induction of sarcomas by mutant IDH2. *Genes Dev*. **27** (18), 1986-98, doi: 10.1101/gad.226753.113 (2013).
30 Kumar, R. *et al.* AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature*. **500** (7460), 89-92, doi: 10.1038/nature12299 (2013).

31 Li, S. *et al.* An optimized algorithm for detecting and annotating regional differential methylation. *BMC Bioinformatics*. **14 Suppl 5**, S10, doi: 10.1186/1471-2105-14-S5-S10 (2013).
32 Patterson, K., Molloy, L., Qu, W., Clark, S. DNA methylation: bisulphite modification and analysis. *Journal of Visualized Experiments. (*56), 3170, doi: 10.3791/3170 (2011).
33 Landan, G. *et al.* Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet*. **44** (11), 1207-14, doi: 10.1038/ng.2442 (2012).
34 Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc*. **7** (12), 2159-70, doi: 10.1038/nprot.2012.137 (2012).
35 Goecks, J., Nekrutenko, A., Taylor, J., Galaxy, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*. **11** (8), R86, doi: 10.1186/gb-2010-11-8-r86 (2010).
36 Dorff, K.C. *et al.* GobyWeb: simplified management and analysis of gene expression and DNA methylation sequencing data. *PLoS One*. **8** (7), e69666, doi: 10.1371/journal.pone.0069666 (2013).
37 Matthias Dodt, J.T.R., Rina Ahmed, Christoph Dieterich. Flexbar − flexible barcode and adapter processing for next-generation sequencing platforms. *MDPI Biology* **1**(3), 895 - 905 (2012).
38 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal, North America*. **17** (1), 10-12, doi: 10.14806/ej.17.1.200 (2011).
39 Bolger, A.M., Lohse, M., Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30** (15), 2114-20, doi: 10.1093/bioinformatics/btu170 (2014).
40 Needleman, S.B., Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. **48** (3), 443 - 453 (1970).
41 Krueger, F., Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. **27** (11), 1571 - 1572, doi: 10.1093/bioinformatics/btr167 (2011).
42 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25** (16), 2078 - 2079, doi: 10.1093/bioinformatics/btp352 (2009).
43 Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res*. **12** (6), 996 - 1006, doi: 10.1101/gr.229102 (2002).
44 Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*. **14** (2), 178-92, doi: 10.1093/bib/bbs017 (2013).
45 Team, R.C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* **ISBN 3-900051-07-0,** http://www.R-project.org/ (2012).
46 Nestor, C., Ruzov, A., Meehan, R., Dunican, D. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-

hydroxymethylcytosine in DNA. *BioTechniques*. **48** (4), 317 - 319, doi: 10.2144/000113403 (2010).
47 Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One*. **5** (1), e8888, doi: 10.1371/journal.pone.0008888 (2010).
48 Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*. **149** (6), 1368 - 1380, doi: 10.1016/j.cell.2012.04.027 (2012).
49 Song, C.X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*. **153** (3), 678-691, doi: 10.1016/j.cell.2013.04.001 (2013).
50 Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. **333** (6047), 1300 - 1303, doi: 10.1126/science.1210597 (2011).
51 Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*. **13** (10), R87, doi: 10.1186/gb-2012-13-10-r87 (2012).
52 Stockwell, P.A., Chatterjee, A., Rodger, E.J., Morison, I.M. DMAP: Differential Methylation Analysis Package for RRBS and WGBS data. *Bioinformatics* **30** (13), 1814-22, doi: 10.1093/bioinformatics/btu126 (2014).
53 Sun, D. *et al.* MOABS: model based analysis of bisulfite sequencing data. *Genome Biol*. **15** (2), R38, doi: 10.1186/gb-2014-15-2-r38 (2014).
54 Bock, C. Analysing and interpreting DNA methylation data. *Nat Rev Genet*. **13** (10), 705 - 719, doi: 10.1038/nrg3273 (2012).
55 Rivera, C.M., Ren, B. Mapping human epigenomes. *Cell*. **155** (1), 39 - 55, doi: 10.1016/j.cell.2013.09.011 (2013).

Figure 1
Click here to download Figure: Figure_1_JoVE_04252014.pdf

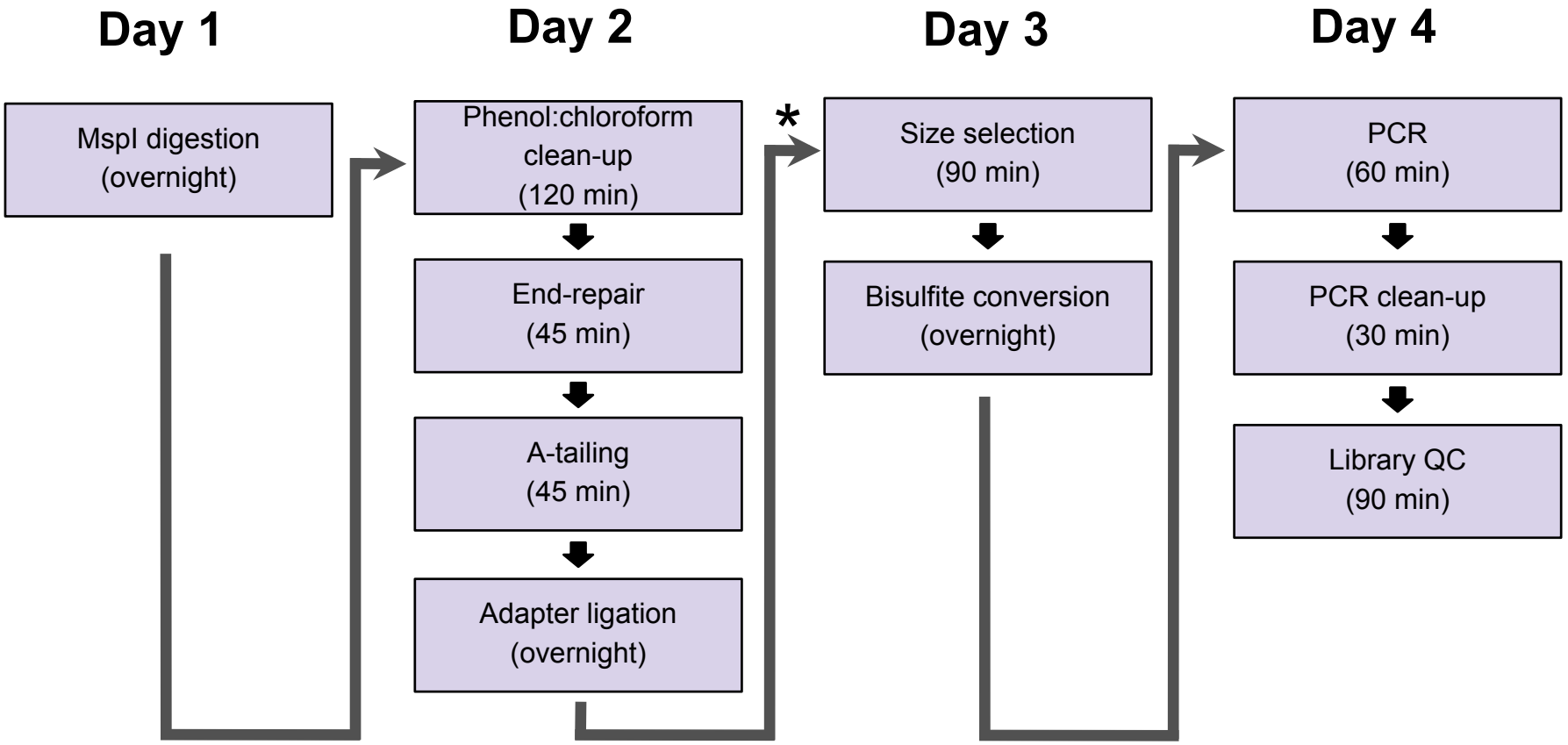
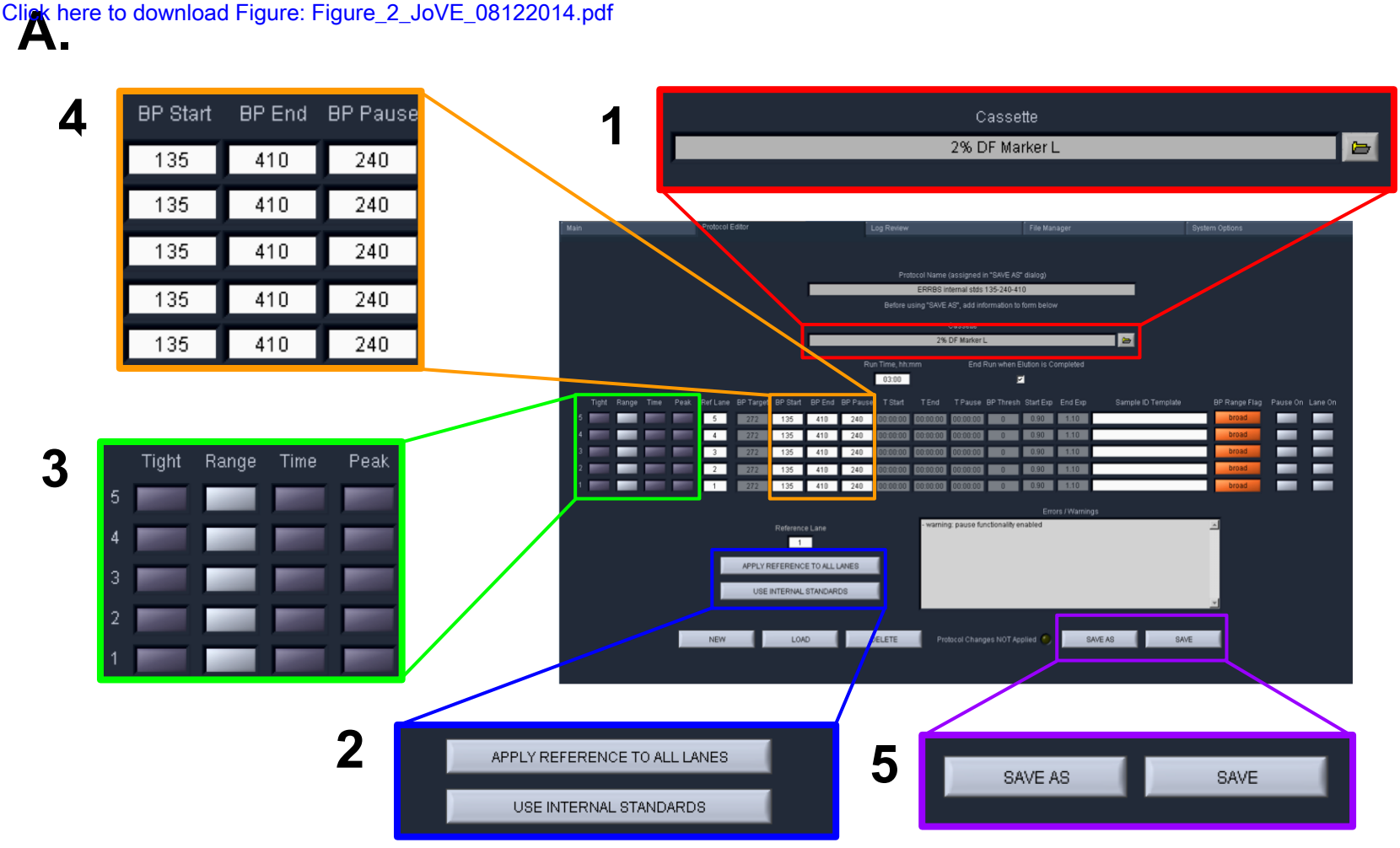
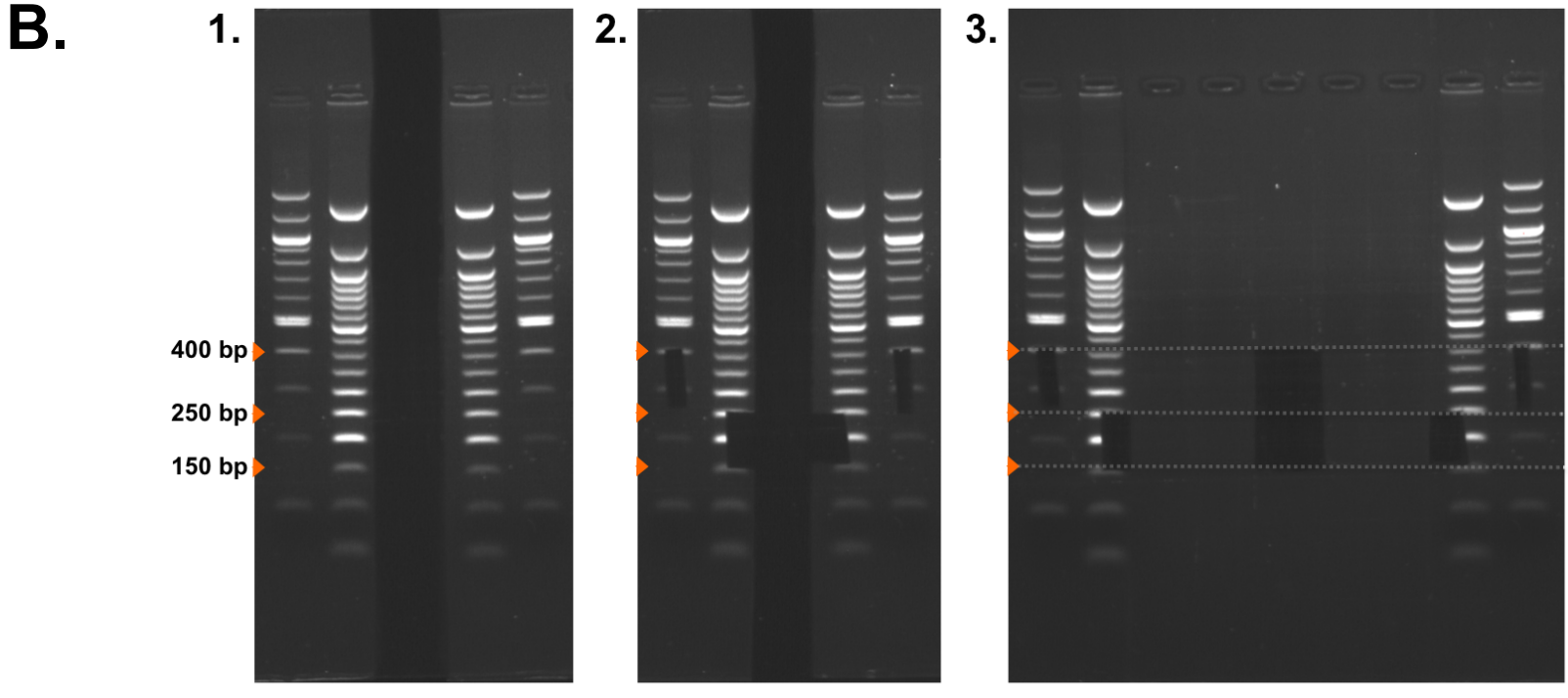


Figure 2

Click here to download Figure: Figure_2_JoVE_08122014.pdf





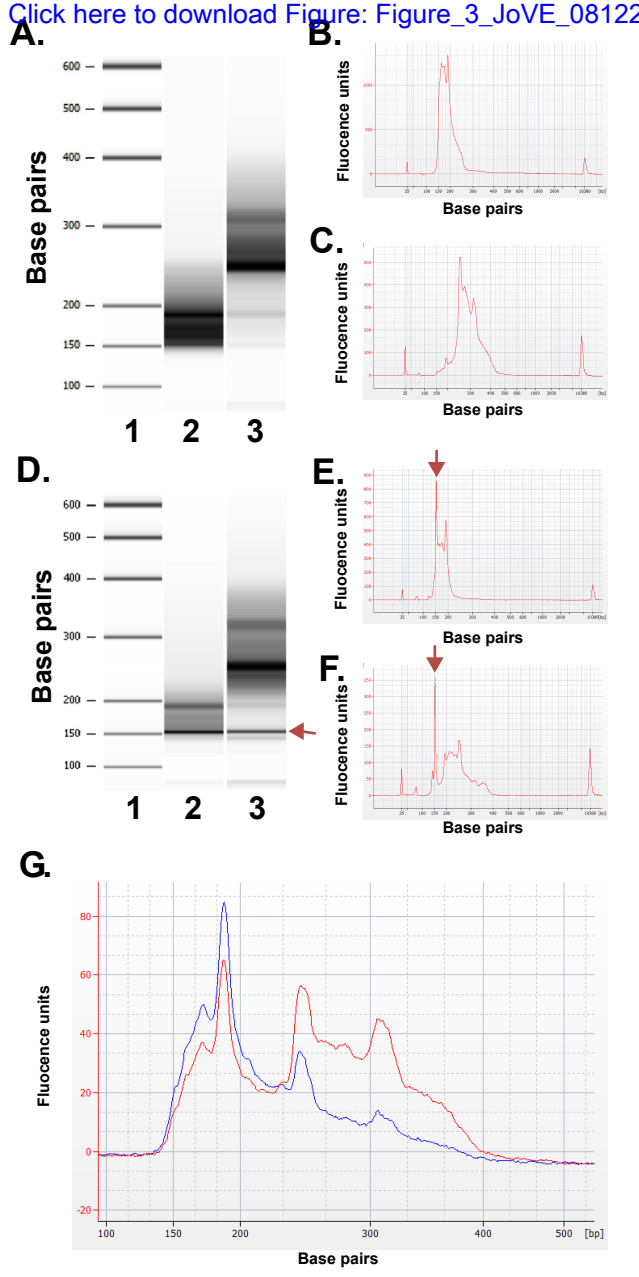
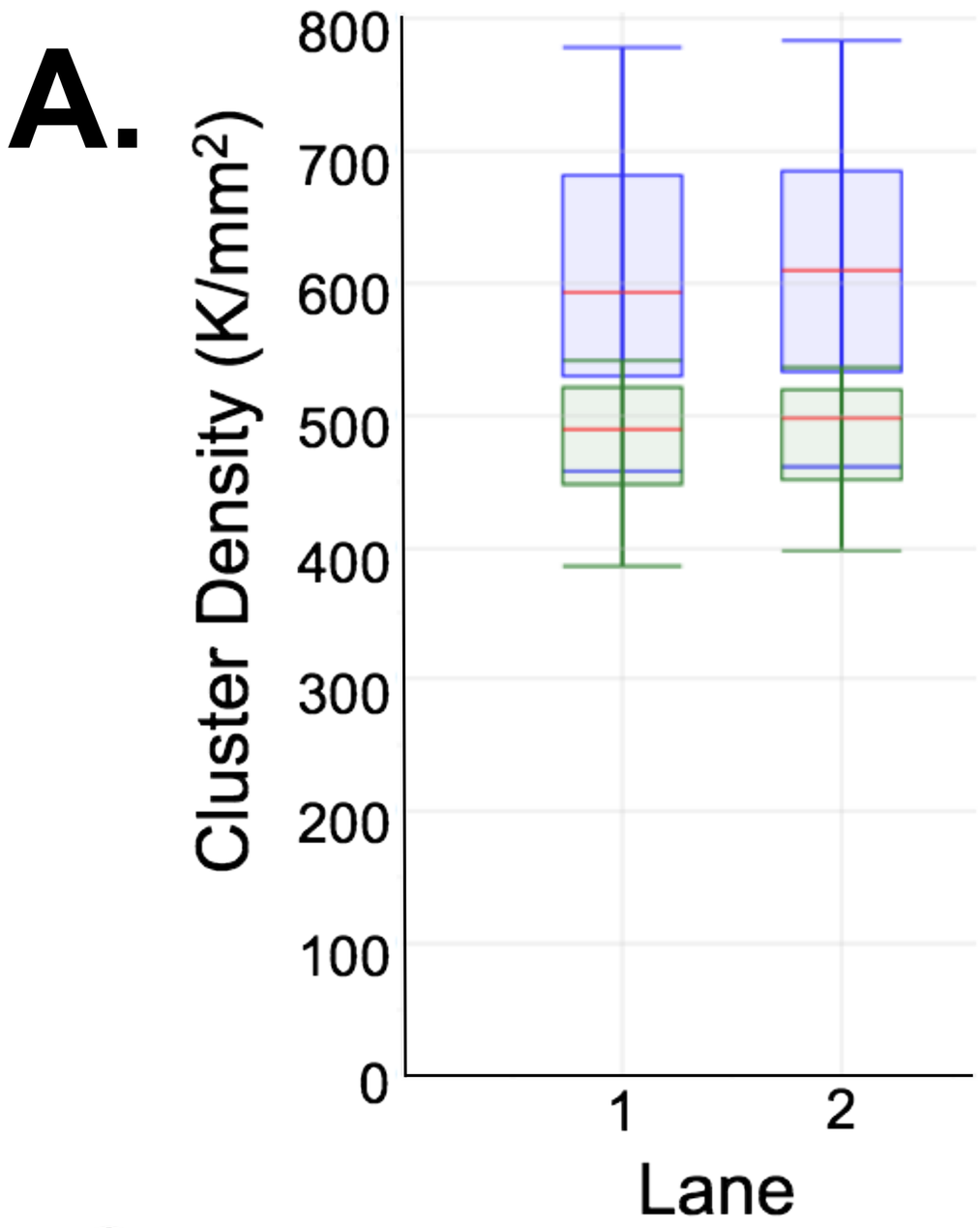
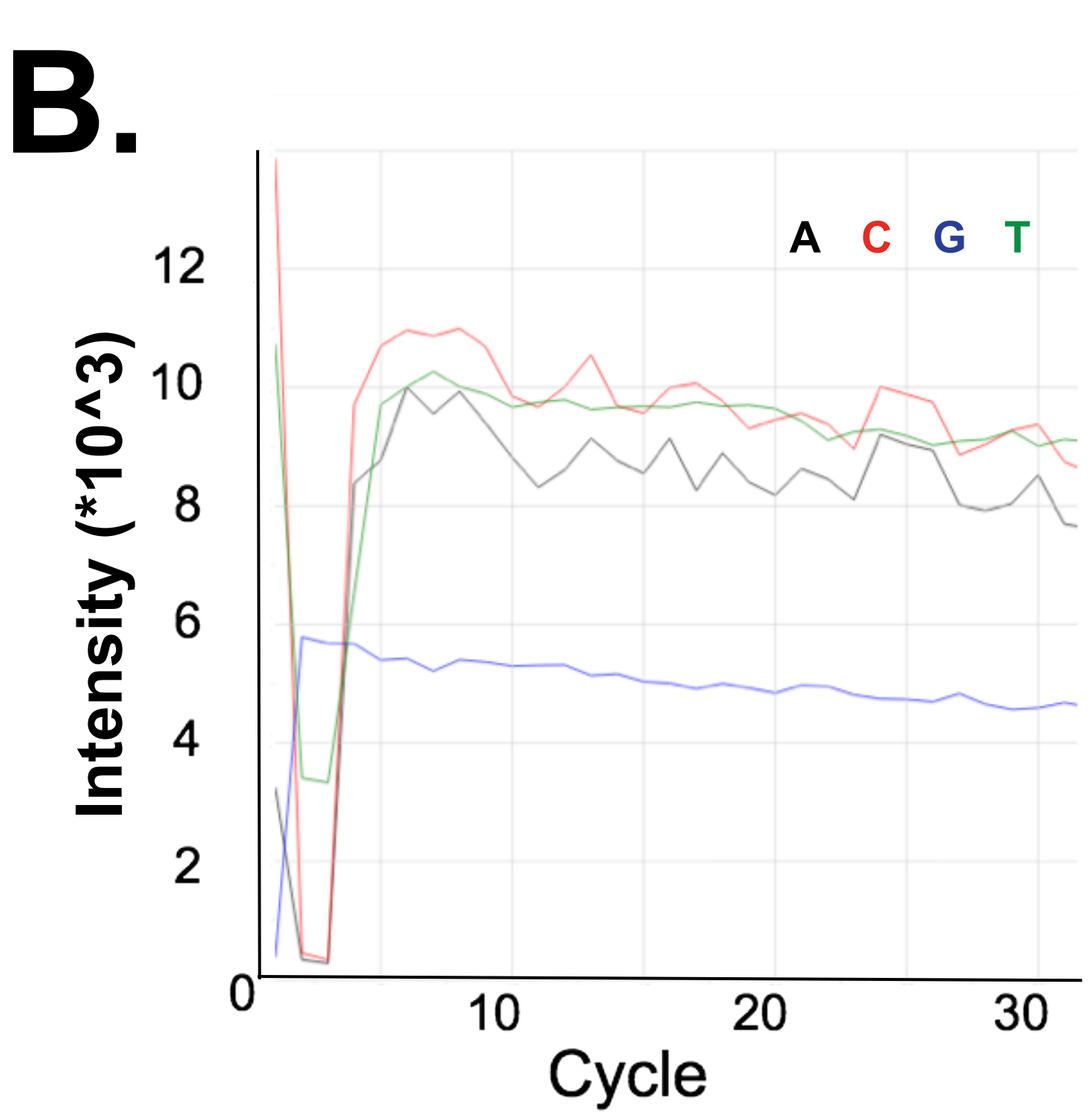
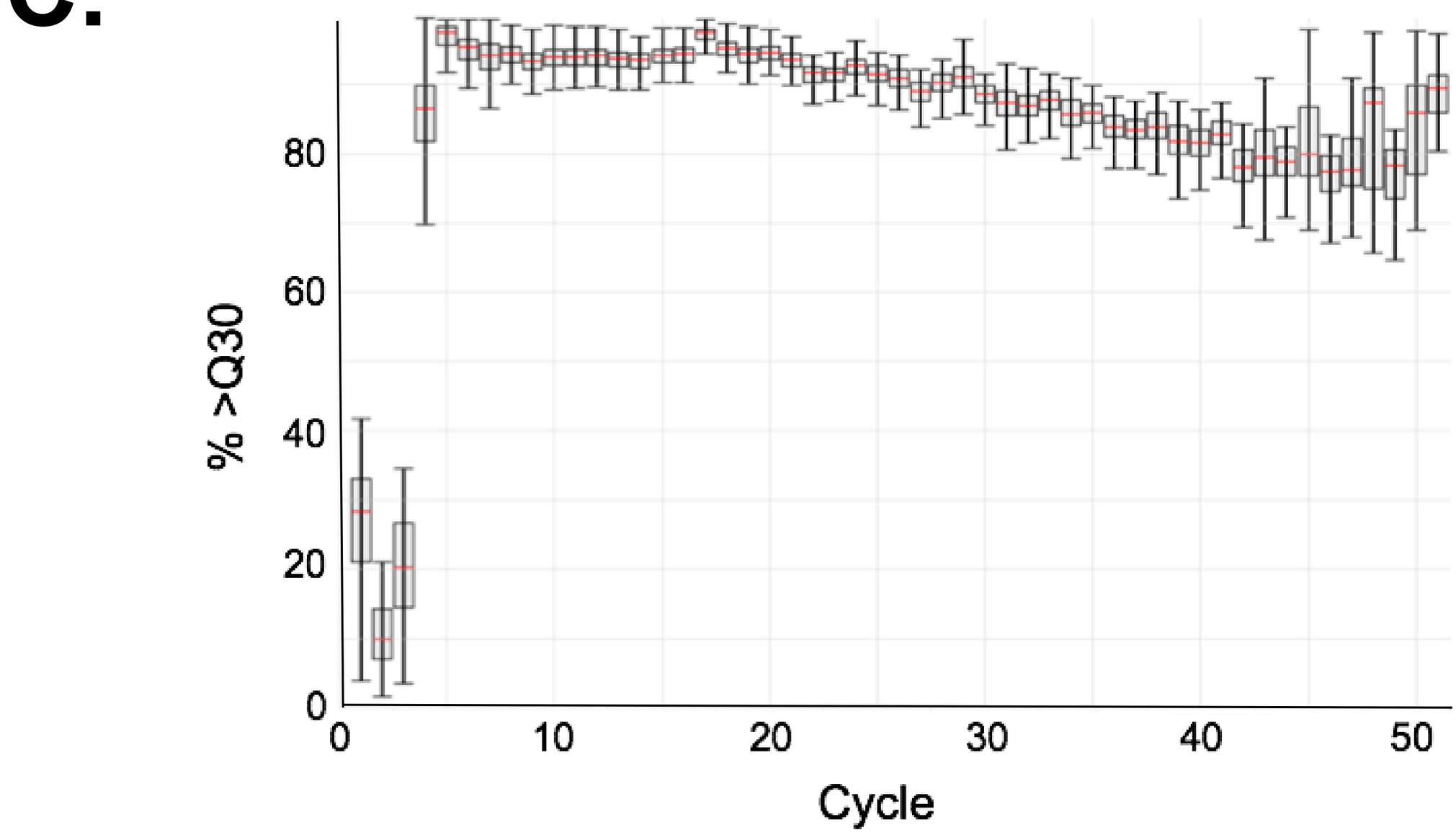


Figure 4
Click here to download Figure: FIgure_4_JoVE_04252014.pdf







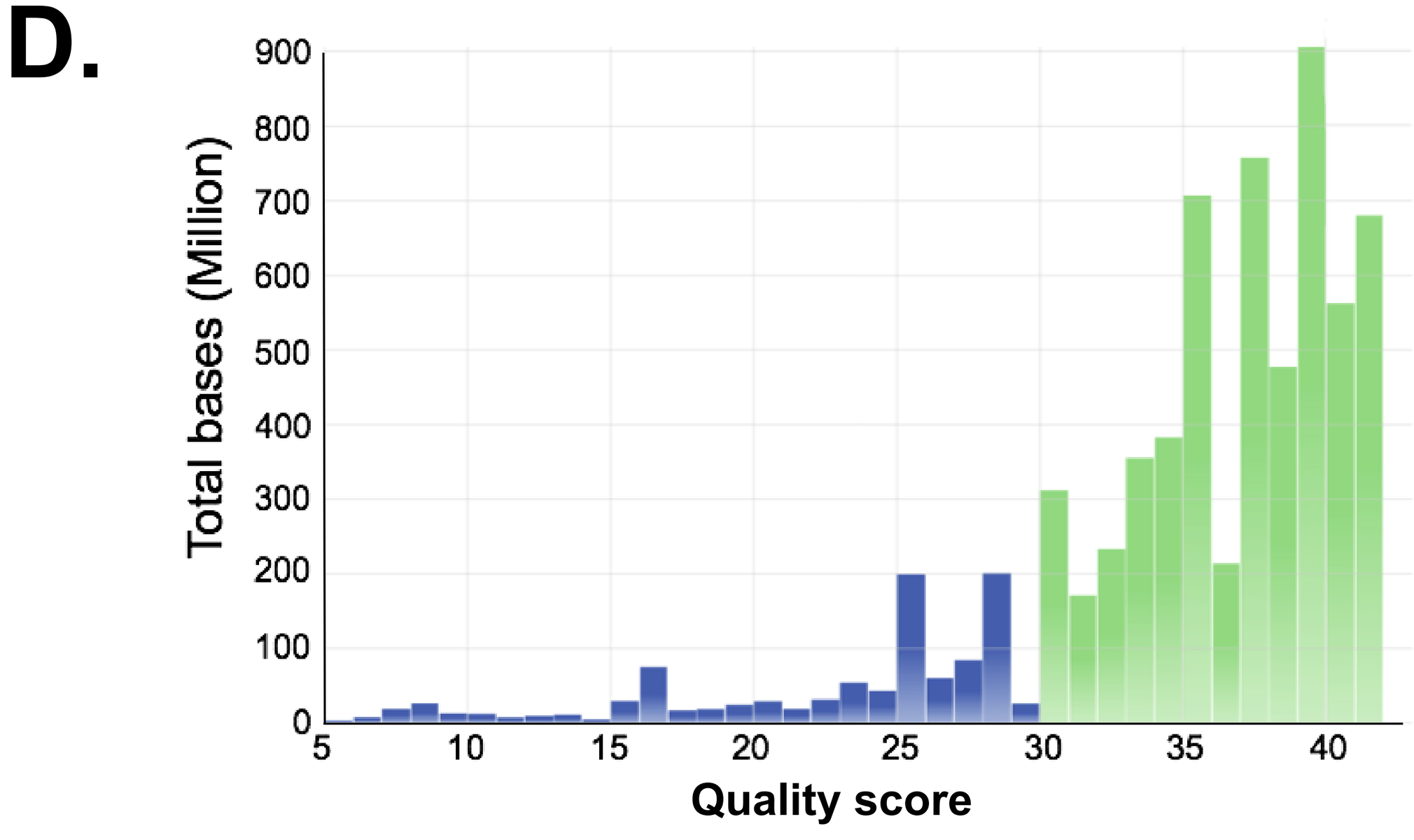
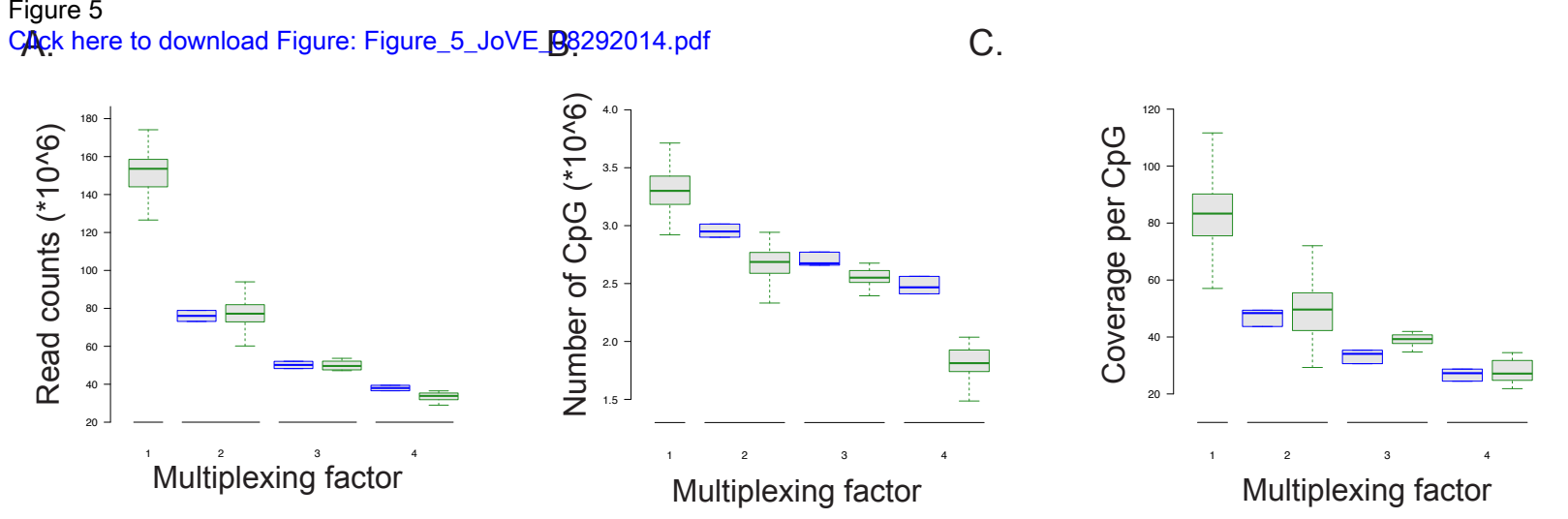


Figure 5

Click here to download Figure: Figure_5_JoVE_28292014.pdf



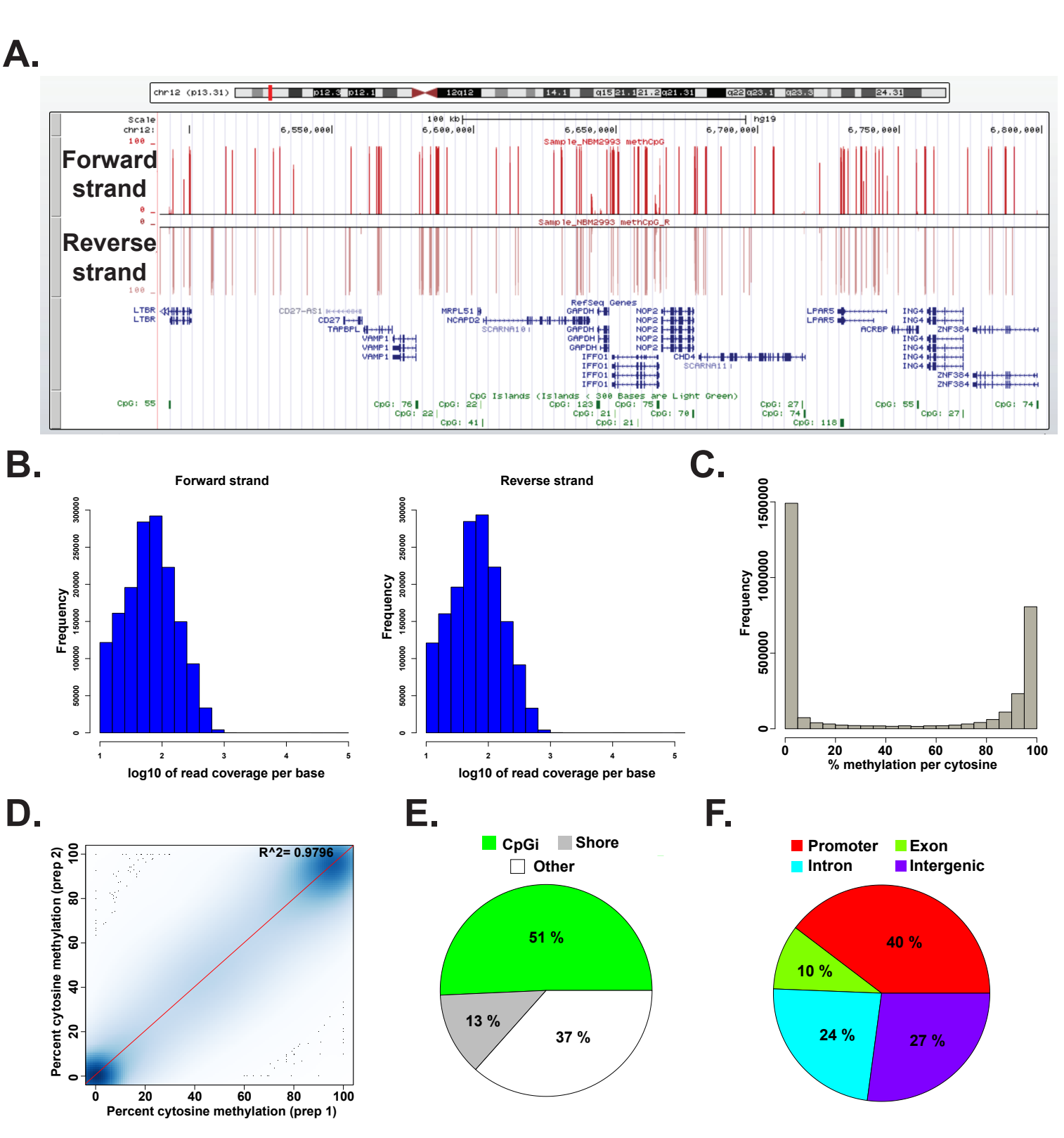

A.
Forward strand
Reverse strand
Sample_NBM2993 methCpG
Sample_NBM2993 methCpG_R
RefSeq Genes
CpG Islands (Islands < 300 Bases are Light Green)
B.
Forward strand
Reverse strand
Frequency
log10 of read coverage per base
C.
Frequency
% methylation per cytosine
D.
R^2= 0.9796
Percent cytosine methylation (prep 2)
Percent cytosine methylation (prep 1)
E.
CpGi
Shore
Other
51 %
13 %
37 %
F.
Promoter
Exon
Intron
Intergenic
40 %
10 %
24 %
27 %

Table 1

Click here to download Table: Table_1_End_repair_08252014.xlsx

Reagent	Volume	Comment
10X T4 DNA Ligase Reaction Buffer	10 µl	
Deoxynucleotide triphosphate (dNTP) Solution Mix	4 µl	mix of 10 mM of each nucleotide
T4 DNA Polymerase	5 µl	3,000 units/milliliter
DNA Polymerase I Large (Klenow) Fragment	1 µl	5,000 units/milliliter
T4 Polynucleotide Kinase	5 µl	10,000 units/milliliter
DNase-free water	45 µl	

Table 2

Click here to download Table: Table_2_A_tailing_08252014.xlsx

Reagent	Volume
10X reaction buffer	5 µl
1 mM 2'-deoxyadenosine 5'-triphosphate (dATP)	10 µl
Klenow Fragment (3'→5' exo-)	3 µl

Comment
for example, NEBuffer 2
5,000 units/milliliter

Table 3

Click here to download Table: Table_3_Ligation_08252014.xlsx

Reagent	Volume
15 µM annealed adapters in DNase-free water	3 µl
10X T4 DNA Ligase Reaction Buffer	5 µl
T4 DNA Ligase	1 µl
DNase-free water	31 µl

Comment
PE adapter 1.0 and PE adapter 2.0; see Table 4 for sequences and reference
2,000,000 units/milliliter

Table 4

Click here to download Table: Table_4_primer_Adapter_08252014.xlsx

Name
PE adapter 1.0
PE adapter 2
hMLH1 Primer I
hMLH1.2 Primer II
PCR PE primer 1.0
PCR PE primer 2.0

Sequence
/5Phos/GAT/iMedC/GGAAGAG/iMedC/GGTT/iMe-dC/AG/iMedC/AGGAATG/iMe-dC//iMedC/GA*G
A/iMe-dC/A/iMe-dC/T/iMedC/TTT/iMe-dC/iMe-dC/iMedC/TA/iMe-dC/A/iMe-dC/GA/iMedC/G/iMe-dC/T/iMe-dC/TT/iMedC//iMe-dC/GAT/iMe-dC/*T
GGAGTGAAGGAGGTTACGGGTAAGT
AAAAACGATAAAACCCTATACCTAATCTATC
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T

Reference	Comments	Protocol step
Gu et al., 2011	* = Phosphorothioate Bond iMe-dC = Int 5-Methyl dC	4.1
Gu et al., 2011	* = Phosphorothioate Bond iMe-dC = Int 5-Methyl dC	4.1
Zymo Research	Universal Methylated Human DNA Standard	6.3
Zymo Research	Universal Methylated Human DNA Standard	6.3
Gu et al., 2011	* = Phosphorothioate Bond	7.2
Gu et al., 2011	* = Phosphorothioate Bond	7.2

Table 5

Click here to download Table: Table_5_PCR_08252014.xlsx

Reagent	Volume
10X FastStart High Fidelity Reaction Buffer with 18 mM magnesium chloride	20 µl
10 mM dNTP Solution Mix	5 µl
25µM PCR PE primer 1.0	4 µl
25µM PCR PE primer 2.0	4 µl
FastStart High Fidelity Enzyme	2 µl
DNase-free water	125 µl

Comment
See Table 4
See Table 4
5 units/µl FastStart Taq DNA Polymerase

Table 6

Click here to download Table: Table6_JoVE_VariousInputs_08252014.xlsx

Protocol step	Reagent/protocol detail	Input DNA amount		
		5-10 ng	25 ng	50 ng
1	MspI enzyme	1µl	2µl	2µl
	MspI digest reaction volum	50	100	100
4	Adapters in ligation reactio	1 µl	2 µl	3 µl
	Ligation reaction volume	20 µl	25 µl	50 µl
5	Size selection protocol	Manual gel only	Pippin Prep or manual gel	Pippin Prep or manual gel
7	PCR primer concentration	25µM	25µM	10µM for 14 cycles; 25 µM for 18 cycles
	Number of PCR cycles	18	18	14-18

Table 7

Click here to download Table: Table7_JoVE_RepResults_08252014.xlsx

Chr	Base	Strand	Coverage	freqC	freqT
chr1	10564	R	366	85.52	14.48
chr1	10571	F	423	91.25	8.75
chr1	10542	F	432	91.2	8.8
chr1	10563	F	429	94.64	5.36
chr1	10572	R	366	96.99	3.01
chr1	10590	R	370	88.11	11.89
chr1	10526	R	350	92	8
chr1	10543	R	368	92.93	7.07
chr1	10525	F	433	91.92	8.08
chr1	10497	F	435	88.74	11.26

Table 8

Click here to download Table: Table8_JoVE_Multiplex_08252014.xlsx

Number of ERRBS libraries per lane	Mean number of uniquely aligned reads	Mean number of CpGs covered	Mean coverage per CpG
1	152231184 +/- 13189678	3183594 +/- 713547	85 +/- 16
2	77680837 +/- 7657058	2674823 +/- 153494	49 +/- 9
3	49938156 +/- 2436865	2552186 +/- 76624	39 +/- 2
4	34457208 +/- 4441686	1814461 +/- 144339	28 +/- 4

Excel Spreadsheet- Table of Materials/Equipment

Click here to download Excel Spreadsheet- Table of Materials/Equipment: MaterialsTable_JoVE_08292014PM_NN.xlsx

Name of Reagent/ Equipment	Company	Catalog Number
MspI	New England Biolabs	R0106M
NEBuffer 2	New England Biolabs	B7002S
Phenol solution	Sigma-Aldrich	P4557
Chloroform	Sigma-Aldrich	C2432
Glycogen	Sigma-Aldrich	G1767
NaOAc	Sigma-Aldrich	S7899
Ethanol	Sigma-Aldrich	E7023
Buffer EB	Qiagen	19086
tris(hydroxymethyl)aminomethane (Tris)	Sigma-Aldrich	T1503
Tris- Ethylenediaminetetraacetic acid (T	Sigma-Aldrich	T9285
T4 DNA Ligase Reaction Buffer	New England Biolabs	B0202S
Deoxynucleotide triphosphate (dNTP) Solution Mix	New England Biolabs	N0447L
T4 DNA Polymerase	New England Biolabs	M0203L
DNA Polymerase I, Large (Klenow) Fragment	New England Biolabs	M0210L
T4 Polynucleotide Kinase	New England Biolabs	M0201L
QIAquick PCR Purification Kit	Qiagen	28104

2'-deoxyadenosine 5'-triphosphate (dAT	Promega	U1201
Klenow Fragment (3'→5' exo-)	New England Biolabs	M0212L
MinElute PCR Purification Kit	Qiagen	28004
T4 DNA Ligase	New England Biolabs	M0202M
Methylation Adapter Oligo Kit	Illumina	ME-100-0010
Agencourt AMPure XP	Beckman Coulter	A63881
Pippin Prep Gel Cassettes, 2% Agarose, dye-free	Sage Science	CDF2010
Certified Low Range Ultra Agarose	Bio-Rad	161-3106
Tris-Borate-EDTA (TBE) buffer	Sigma-Aldrich	T4415
Ethidium bromide solution	Sigma-Aldrich	E1510
50 bp DNA Ladder	NEB	N3236S
100 bp DNA Ladder	NEB	N3231S
Gel Loading Dye, Orange (6X)	NEB	B7022S
Scalpel Blade No. 11	Fisher Scientific	3120030
QIAquick Gel Extraction Kit	Qiagen	28704
EZ DNA Methylation Kit	Zymo Research	D5001
EZ DNA Methylation-Lightning Kit	Zymo Research	D5030
Universal Methylated Human DNA Standard	Zymo Research	D5011
FastStart High Fidelity PCR System	Roche	03553426001

Qubit dsDNA High Sensitivity Assay Kit	Life Technologies	Q32854
DynaMag-2 Magnet	Life Technologies	12321D
High Sensitivity DNA Kit	Agilent Technologies	5067-4626
2100 Bioanalyzer	Agilent Technologies	
PhiX Control v3	Illumina	FC-110-3001
HiSeq 2500	Illumina	
Pippin Prep	Sage Science	
Qubit 2.0 Fluorometer	Life Technologies	Q32872
TruSeq SR Cluster Kit v3-cBot-HS	Illumina	GD-401-3001
TruSeq SBS Kit v3-HS	Illumina	FC-401-3002
TruSeq RNA Sample prep	Illumina	RS-122-2001
Microcentrifuge		
Vortex Mixer		
Dry Block Heater		
Thermal Cycler		
Water Bath		
Gel electrophoresis system		
Electrophoresis power supply		
Gel doc		
UV or blue light transilluminator		

Comments/Description
100,000 units/ml
Reaction buffer for MspI enzyme; protocol step 1.2
Equilibrated with 10 mM Tris HCl, pH 8.0; see safety and handling instructions at http://www.sigmaaldrich.com/catalog/product/sigma/p4557
See safety and handling instructions at http://www.sigmaaldrich.com/catalog/product/sial/c2432
19-22 mg/ml
3M pH 5.2
200 proof, for molecular biology
10 mM Tris-Cl, pH 8.5
prepare a 1M pH 8.5 solution
Dilute to 1X buffer solution per manufacturer's recommendations
10X concentration
10 mM each nucleotide
3,000 units/ml
5,000 units/ml
10,000 units/ml
Used for DNA product purification in protocol step 2.3

100 mM
5,000 units/ml
Used for DNA product purification in protocol step 3.3
2,000,000 units/ml
Used in protocol sections that implement magnetic bead purification steps (steps 4.3 and 8.2). Equilibrate to room temperature before use
with internal standards
10 mg/ml
Used in protocol step 6.2
Alternative for step 6.2
Used as bisulfite conversion control

A fluorescence-based DNA quantitation assay; used in protocol steps 1.1, 9.1 and 10.1
Barcoded adapters used for multiplexing libraries; See Supplemental file for multiplexing protocol

URL
https://www.neb.com/products/r0106-mspi
https://www.neb.com/products/b7002-nebuffer-2
http://www.sigmaaldrich.com/catalog/product/sigma/p4557
http://www.sigmaaldrich.com/catalog/product/sial/c2432
http://www.sigmaaldrich.com/catalog/product/sigma/g1767
http://www.sigmaaldrich.com/catalog/product/sigma/s7899
http://www.sigmaaldrich.com/catalog/product/sial/e7023
http://www.qiagen.com/products/catalog/lab-essentials-and-accessories/buffer-eb
http://www.sigmaaldrich.com/catalog/product/sigma/t1503
http://www.sigmaaldrich.com/catalog/product/sigma/t9285
https://www.neb.com/products/b0202-t4-dna-ligase-reaction-buffer
https://www.neb.com/products/n0447-deoxynucleotide-dntp-solution-mix
https://www.neb.com/products/m0203-t4-dna-polymerase
https://www.neb.com/products/m0210-dna-polymerase-i-large-klenow-fragment
https://www.neb.com/products/m0201-t4-polynucleotide-kinase
http://www.qiagen.com/products/catalog/sample-technologies/dna-sample-technologies/dna-cleanup/qiaquick-pcr-purification-kit

http://www.promega.com/products/pcr/routine-pcr/deoxynucleotide-triphosphates-_dntps_/
https://www.neb.com/products/m0212-klenow-fragment-3-5-exo
http://www.qiagen.com/products/catalog/sample-technologies/dna-sample-technologies/dna-cleanup/minelute-pcr-purification-kit
https://www.neb.com/products/m0202-t4-dna-ligase
https://www.beckmancoulter.com/wsrportal/wsrportal/wsr/research-and-discovery/products-and-services/nucleic-acid-sample-preparation/agencourt-ampure-xp-pcr-purification/index.htm?i=A63882#2/10//0/25/1/0/asc/2/A63882///0/1//0/
http://store.sagescience.com/s.nl/it.A/id.1036/.f
http://www.bio-rad.com/en-us/sku/161-3106-certified-low-range-ultra-agarose
http://www.sigmaaldrich.com/catalog/product/sigma/t4415
http://www.sigmaaldrich.com/catalog/product/sigma/e1510
https://www.neb.com/products/n3236-50-bp-dna-ladder
https://www.neb.com/products/n3231-100-bp-dna-ladder
https://www.neb.com/products/b7022-gel-loading-dye-orange-6x
http://www.fishersci.com/ecomm/servlet/fsproductdetail?position=content&tab=Items&productId=11876776
http://www.qiagen.com/products/catalog/sample-technologies/dna-sample-technologies/dna-cleanup/qiaquick-gel-extraction-kit
http://www.zymoresearch.com/epigenetics/dna-methylation/bisulfite-conversion/ez-dna-methylation-kits/ez-dna-methylation-kit
http://www.zymoresearch.com/epigenetics/dna-methylation/bisulfite-conversion/ez-dna-methylation-lightning-kit
http://www.zymoresearch.com/epigenetics/dna-methylation/methylated-dna-standards/universal-methylated-human-dna-standard
http://lifescience.roche.com/shop/products/faststart-high-fidelity-pcr-system#tab-0

http://www.lifetechnologies.com/order/catalog/product/Q32854
https://www.lifetechnologies.com/order/catalog/product/12321D
http://www.genomics.agilent.com/en/Bioanalyzer-DNA-RNA-Kits/High-Sensitivity-DNA-Analysis-Kits/?cid=AG-PT-105&tabId=AG-PR-1069
http://www.genomics.agilent.com/en/Bioanalyzer-System/2100-Bioanalyzer-Instruments/?cid=AG-PT-106&tabId=AG-PR-1001
http://www.illumina.com/products/phix_control_v3.ilmn
http://www.illumina.com/systems/hiseq_2500_1500.ilmn
http://www.sagescience.com/products/pippin-prep/

http://www.lifetechnologies.com/order/catalog/product/Q32872

http://www.illumina.com/products/truseq_sr_cluster_kit_v3-cbot-hs.ilmn
http://www.illumina.com/products/truseq_sbs_kit_v3-hs.ilmn
http:/http://www.illumina.com/products/truseq-rna-access-kit.ilmn

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Enhanced Reduced Representation Bisulfite Sequencing for assessment of DNA methylation at base pair resolution

Author(s): Francine E. Garrett-Bakelman, Caroline K. Sheridan, Thadeous J. Kacmarczyk, Jennifer Ishii, Alicia Alonso, Maria E. Figueroa, Christopher E. Mason, Ari M. Melnick

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via: ☐ Standard Access ☒ Open Access

Item 2 (check one box):

☒ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name: Francine E. Garrett-Bakelman

Department: Department of Medicine

Institution: Weill Cornell Medical College

Article Title: Enhanced Reduced Representation Bisulfite Sequencing for assessment of DNA methylation at base pair resolution

Signature:  Date: 4/24/14

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

Response to reviewers
Manuscript JoVE52246R2
'Enhanced Reduced Representation Bisulfite Sequencing for assessment of DNA methylation at base pair resolution'

We appreciate the opportunity to revise our manuscript and are hereby submitting revised documents along with responses to reviewers. Editorial and reviewer comments are in italic bold font below and responses to comments follow each respectively.

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.
Thank you for the opportunity to review the manuscript for any errors. All authors have thoroughly proofread the manuscript.

2. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents.
For example: All Cells, LLC, etc.
We recognize JoVE's restrictions on the publication of commercial language and the AllCells terminology has been removed

3. Please double-check the in text references to Protocol steps for accuracy.
Thank you for the opportunity to review the manuscript to ensure reference list accuracy. All authors have reviewed this information

4. Please place all URLs in the Table of Materials and Equipment.
All pertinent URLs have been included in the Table of Materials and Equipment. For large equipment that is commonly used in laboratories, no specific URL was included.

5. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted.

6. Please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol. The highlighted steps should form a cohesive

narrative with a logical flow from one highlighted step to the next. Remember that non-highlighted Protocol steps will remain in the manuscript, and therefore will still be available to the reader.
Thank you for the opportunity to review the manuscript to ensure all protocol steps are properly highlighted. Revisions to the sections annotated for the potential video have been included in the updated version.

Reviewers' comments:

Reviewer #1:
This article titled "Enhanced Reduced Representation Bisulfite Sequencing for assessment of DNA methylation at base pair resolution" is a clearly written article that describes the steps involved in generating the high-throughput sequencing libraries for the researchers interested in cytosine methylation patterns at CpG rich regions at a reduced cost. The steps of the protocol for ERRBS are written with clarity giving details about the reagents and the alternatives available. This article is well within the scope of the journal and the details are sufficient for this type of article. In this article, the authors have started by introducing the role of cytosine methylation in the genome, methods in use to analyze the cytosine methylation and the advantage of ERRBS over the available methods. All the steps in the protocol were explained with enough details that will enable reproducibility. Authors have also extended the article by briefing about the methods used to analyze the generated data. Figures are representative, simple and of good quality. Figure legends explain the figures very well.
We thank reviewer one for his/her thorough input. Below are comments to address each point raised.

However, there are a few points to clarify and typos to correct:
All through the protocol, the authors have treated the two fragments sizes (135-240 and 240-410 bp) separately and subjected them to similar reactions and treatments. Finally they mixed the two fragment sizes. Authors may clarify why they were treated the fragment sizes separately and not as a single fraction (135-410). What is the advantage in treating them separately while authors are not excluding any intermediate fragment sizes? Is this specific to Pippin Prep?
The rationale for the division of the library product into two fragments is to ensure efficient library product amplification during the PCR step (protocol number 7). Splitting the fractions for amplification and then combining them in equimolar amounts to make the final library for sequencing results in a more equal distribution of fragments across the entire size range. The bioanalyzer traces below illustrate the difference between a library created from a single 150-400 bp fraction (blue trace) and a library created from two fractions (red trace). The library created from a single fraction skews toward smaller fragments while the two-fraction library has a more even distribution of fragments.

It is likely beneficial to the users of this protocol to have a manual gel-based size selection protocol, as many (most) may not have access to a Pippin Prep. Although one is referenced, this would provide a single protocol for ERRBS.

We thank the reviewer for this recommendation. Protocol steps for a manual gel-based size selection have been included in the revised manuscript (protocol section 5.2).

Lines 538-543 give a good description of sequenced read depth to mapped reads to # of CpGs obtained. However, the authors could aid the readers/users by providing a barchart/histogram/graph that shows these numbers in 10 or 20 million sequenced read increments; i.e. for each read depth increment how many reads mapped, and how many CpGs are recovered or gained by going deeper. This will allow the user to judge how deep to sequence. The example given is for a full lane, which seems rare for RRBS experiments. If a full lane isn't needed for ERRBS, as is the case for RRBS, then multiplexing should be noted.

We thank the reviewer for this recommendation. We have added summarized results from multiplexed sequencing runs in the manuscript. We have added a figure (Figure 5) representing this information in comparison to expected data from sequencing a single ERRBS library per lane and simulated downsampling results from three representative sequencing runs (sampling 5 times for each comparison). We have also added Table 4, which includes average results from multiplexing ERRBS libraries that we have performed in the past year. Commentary about the expected results in also included in the results section.

The authors may consider discussing alternatively spiking in sonicated lambda DNA for the bisulfite conversion efficiency with the caveat that it is only detected after sequencing the library not prior.

We thank the reviewer for this recommendation. We agree that such a control is beneficial to include with the disadvantage of not being able to determine conversion efficiency until after sequencing. Should the conversion efficiency be suboptimal, this process would pose a high financial cost and we recommend the use of a parallel control that can be independently validated as described in the protocol. We have revised the manuscript to include this information for user's choice as a note in protocol section 6 and have referenced a publication which describes such an approach[1].

At lines 465-474, the authors have specified the usage of "FAR" software referring to an article about "Flexbar" software (Ref 28) for trimming the reads. I noticed that "FAR" is now defunct and not available for downloads. Authors should clarify the name of the software and its reference (if they wish to name it) and its reference.

The reviewer is correct. Flexbar is the current name of the FAR software, and we have updated the text (protocol step 11.3) to reflect this information inclusive of the current reference for this tool. Any software tool that removes adapter sequences could be used (cutadapt, trimmomatic, etc.) in this step of the analysis and we have included additional examples in the manuscript for the readers' consideration.

At lines 493-514, Bismark is a good tool but is under continuous development with additional scripts available for format conversion and also enables usage of multiple aligners (Bowtie and Bowtie 2 for ungapped and gapped alignments). So, either authors may update the text to refer Bismark's perl scripts or generalize the word "PERL script" to "custom scripts".

We thank the reviewer for this suggestion. We have updated the text to recommend the use of custom scripts when appropriate.

Figure 5D is misleading without contrasting the R squared value of different biological samples. RRBS enriches for CGs where the majority are largely 100% methylated or non-methylated; therefore it is likely that even two different cell types will have an equally high correlation.

We thank the reviewer for this comment. We have performed correlation analyses of CpGs covered and their methylation levels on various sample types and in all cases the R^2 values were below that reported for the technical replica included in this manuscript. We have included a few examples below:

Sample 1	Sample 2	R^2 value
IMR90	HCT116	0.4675
AML3	CB	0.8385

AML3	NB	0.8416
AML1	AML2	0.7665
AML1	AML3	0.8647
AML2	AML3	0.7686

IMR90 = normal human lung fibroblast cell line
HCT116 = human colorectal cancer cell line
AML = Acute Myeloid Leukemia samples (1-3)
CB = centrocyte B cells isolated from tonsils
NB = Naïve B cells isolated from tonsils

We have modified the text to indicate R^2 value ranges expected between technical versus biological samples as above-noted. Should the reviewers deem additional figures necessary, we can include those as well.

@line 164-168: pH of the Tris-Cl is missing. Since pH of the buffer is crucial and determines the phase separation of DNA and RNA, it is worth mentioning the pH of the buffer.
Thank you for the reviewer's attention to this omission. We agree that buffer pH's are essential to successful protocol implementation and the information has been included in the revised manuscript.

@line 167: following Tris- Ethylenediaminetetraacetic acid "(TE)" could be shown
We have included this edit in our updated manuscript

@line 178: Alcohol precipitation is a method that depends on the concentration of the alcohol and sodium acetate; therefore, the authors may consider rephrasing the amount of alcohol used (750 microlitres be changed to 2.5 volumes).
Thank you for the recommendation; we have revised the protocol to ensure clarity and have added volumes to the solution quantities already noted.

@line 244: "unicorporated" is misspelled.
Thank you for the reviewer's comment. We have revised the manuscript accordingly

@line 511-514: The recent versions of Bismark generate the SAM formatted output by default. No need of any further scripts to process Bismark's output. Authors may generalize about the possible ways to visualize the methylation data (there are other alternative ways to visualize the data in addition to wiggle format).
We thank the reviewer for this comment. We use a previous version of Bismark that does not generate SAM formatted output but instead provides a bismark2SAM_v5_xm.pl to convert Bismark alignments to SAM format which we then convert to BAM format using SAMTOOLS. In the updated manuscript, we have included additional viewing options to wiggle track formats, including

bedGraph format, the UCSC genome browser and IGV.

Reviewer #2:
Garrett- Bakelman et al have described a high resolution genomic methylation analysis platform (e-RRBS) protocol that will benefit epigenetic researchers especially those studying primary tissue samples with low input DNA.

The article is well written and organized and the companion video will help enhance the visualization of the protocol.
We thank reviewer two for his/her thorough input. Below are comments to address each point raised.

Major concerns:
1. Steps 10 and 11 have not been included in the visualization - these will be important especially to convey key aspects of the data analysis (step 11)
We thank the reviewer for this recommendation. Due to the length of the protocol and JoVE's guidelines for highlighted segment inclusions, we will be unable to incorporate step 10 and 11 into the visualization process. We agree that these steps are essential to successful implementation of the protocol and hence have added additional information in them. We have revised our manuscript to include further details on the library pooling protocol (protocol step 10). Furthermore, we have expanded our supplemental code file to include a full list of detailed commands for data alignment.

Minor concerns:
1. Use of a high complexity library lane as a control for cluster density during sequencing is mentioned on page 15. Can the authors add 1-2 sentences on how this is used?
We thank the reviewer for this recommendation and we agree that further explanation would be helpful. We have modified the manuscript in the discussion to clarify the recommendation to use a high complexity library lane (PhiX control or other library type) as a dedicated independent control lane. When designated as a control lane for the respective sequencing run, it can serve as the basis for the matrix generation which is utilized during the first four bases of sequencing to detect cluster positions. The higher quality reads captured will raise the mean coverage per CpG site by 5.2 (n=4).

2. Add in discussion whether the technique can be adapted for multiplexing?
We thank the reviewer for the recommendation. We have expanded the manuscript to include further detail on how the protocol could be adapted for multiplexing (protocol step 12). We have also added figure 5 which highlights expected outcomes from standard and multiplexed approaches.

Reviewer #3:

Editor's Note: *Please note that the reviewers raised some significant concerns regarding your method and your manuscript. Please thoroughly address each concern by revising the manuscript or addressing the comment in your rebuttal letter.*

Manuscript Summary:
This protocol describes what the authors consider an improvement of the original RRBS protocol and which they called "enhanced RRBS". However, the changes compared to the original protocol are incremental in nature and do not justify the new name, especially because the improvements in coverage come at the cost of significantly increased costs per sample and therefore constitute a different trade-off rather than a dominant advance.

Furthermore, the comparison between the original RRBS protocol and the "enhanced RRBS" protocol is no longer relevant because the original RRBS protocol has been superseded by much-improved versions in many labs, which provide various improvements that the described "enhanced RRBS" protocol is lacking.
We thank reviewer three for a thorough review of our manuscript. The manuscript submitted details the current protocol we use based on our previous publication (Akalin and Garrett-Bakelman, et al. PLoS Genetics **8**(6) in 2012) with the goal of providing guidance and sufficient detail for setting up and implementing the protocol in a laboratory novice to this technique. We have addressed each concern raised below.

Compared to the current state-of-the-art, the most severe flaws of the presented protocol are as follows:

1. Low throughput: The protocol contains several steps that are inherently low throughput, which makes it impossible to effectively process large cohorts*.*
We agree that the ERRBS protocol as described is limited to the preparation of 10 - 15 samples per week by a single individual. In our experience, the vast majority of projects profiled include up to 30 samples. The protocol could be converted to a high throughput protocol if the user or lab implementing it has an automated next generation library preparation workstation. However we do not think this is feasible in most labs and we have used the described approach successfully and reproducibly in over 2000 samples.

2. High cost: The protocol is clearly not optimized for cost and lacks even basic features such as support for sample barcoding and multiplex sequencing.
We thank the reviewer for this comment. We routinely perform our protocol to acquire data with high coverage per CpG site in order to perform analysis that requires high coverage depth (for example to detect epigenetic polymorphisms

as described in Landan, et al.[2]). While we agree that barcoding and multiplexing ERRBS will yield a lower cost per sample protocol, this will result in a reduction in the number of CpGs covered and the coverage at each CpG site represented in the data. We have revised our manuscript to include further details on how the protocol can be modified for a multiplexed sequencing approach (protocol step 12) and have included data on the number of CpGs and depth of coverage anticipated from such an approach for the readers' consideration (Figure 5 and table 4).

3. Infeasible for most labs: The protocol uses specialized equipment for size selection that is not only a severe throughput bottleneck but also unavailable in most labs.

We thank the reviewer for this comment. We have incorporated a section in the protocol describing a gel extraction approach for size selection, which should be feasible in all laboratories (step 5.2). Unfortunately unless the protocol is automated, this step is a bottleneck, as it is in many other next generation sequencing library preparation protocols.

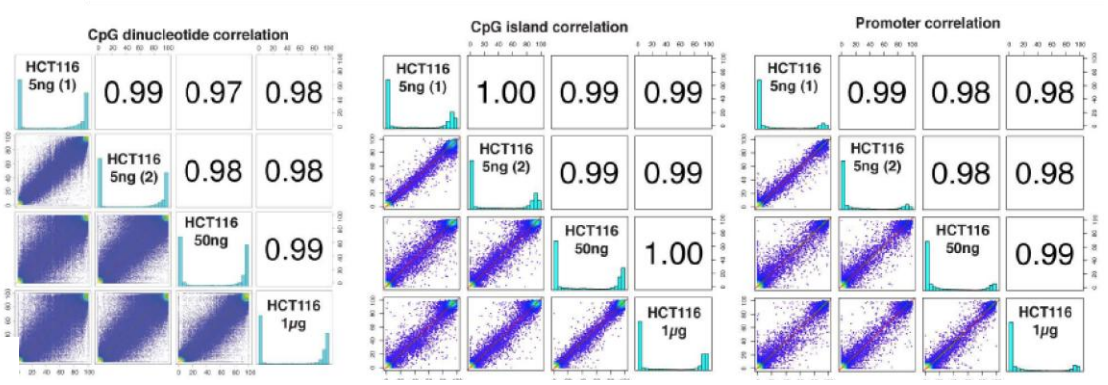
4. No proof of robustness: I would expect that the authors summarize their experience with the protocol for at least a couple of dozen (or better: hundreds of) samples run in their lab over the last years, such that the reader gets an impression of the protocol's robustness.

We thank the reviewer for his/her comment about protocol robustness. We had indicated in the representative results section of the manuscript our experience in terms of data quality and quantity from human samples sequenced at 50bp single end reads in the past few years (cluster densities, pass filter rates, number of uniquely aligned reads per lane, alignment efficiency, the number of CpGs covered per sample and the average coverage at each CpG covered). Should there be any additional information the reviewer deems appropriate for inclusion, we would be happy to do so.

Additional Concerns:

1. The protocol contains far too many cleanup steps that require a large amount of input DNA. Although the authors claim that the procedure would work even down to 5ng of input DNA they never actually show data that support this claim.

We thank the reviewer for this comment. In our PLoS Genetics paper (Akalin and Garrett-Bakelman, et al. 2012), we had performed a titration of input DNA (5, 50 and 1000ng) which revealed high reproducibility of our current ERRBS protocol when data was examined for all CpGs covered as well as CpGs found in CpG islands and promoters (pearson correlation between technical replicas of ERRBS using 5, 50 or 1000 ng genomic DNA from the HCT116 cell line). This has been referenced in this manuscript.



2. The described bisulfite conversion protocol has been superseded by newer and more efficient protocols.

We agree that several newer kits for bisulfite conversion have been brought to the market however we have found that the kit we use is highly reproducible. In our hands, the bisulfite conversion rates observed are consistently higher than those reported to be guaranteed by kits offered by Zymo research and Epigentek (99 – 99.5%), with an average conversion rate from 390 sequencing runs of 99.85 +/-0.04 %. We have added alternative reagents to the Table of Materials and Equipment which may be used should the user prefer.

3. The bisulfite conversion controls are added too late. Why don't the authors spike them in at the very beginning (for end repair, A-tailing and ligation)? This would enable them to control individual molecules on the HiSeq sequencer and not just a mixture of individual signals on the Sanger sequencer. Furthermore, no over-conversion controls are used.

We agree that a spike-in control could be beneficial to include. However a disadvantage compared to the control detailed in the protocol is not being able to determine conversion efficiency until after sequencing. Should the conversion efficiency be suboptimal, this process would result in low quality data due to inefficient conversion and pose a high financial cost that could have been avoided. We recommend the use of a parallel control that can be independently validated as described. We have revised the manuscript to include the option of a spike-in control for the user's choice (Note in protocol step 6). While we do not routinely test for over conversion, thorough review of this concern with the research division at ZymoResearch's revealed that over-conversion is estimated to be sub-percent and that the impact of over-conversion on data quality would not be significant unless the bisulfite treatment would be extended for an additional four hours than currently used. Finally, review of select papers published about RRBS and other related protocols does not reveal routine use of an over-conversion control in published protocols to date (for example [1,3-5]).

4. Running 18 cycles by default can lead to an amplification bias in some cases. A qPCR should be performed first, to calculate the ideal cycle

number for enrichment.
We agree that a qPCR step for library quantitation prior to PCR amplification could be considered for inclusion in the protocol, however we routinely do not implement it. Based on our assessments, introducing a qPCR step will not result in a change in the protocol proposed and 18 cycles of PCR yield sufficient material for sequencing from 5-50ng starting materials (typical yields are: 25 µl of 30 nM for lower fraction and 25µl of 8 nM for the higher fraction from 5ng starting materials and 50 µl of 300 nM for lower fraction and 50 µl of 20 nM for higher fraction from 50 ng starting materials). This protocol is continuously under improvement and qPCR quantitation is one aspect we are exploring for further development. We have included a comment in the discussion raising the possibility of using qPCR for further PCR optimization.

5. For a technique like RRBS where there is no possibility to discriminate for PCR duplicates, 18 PCR cycles are already the upper limit that should be applied to a library during enrichment. I don't see how an efficient library should be generated with 5ng of starting material when there is the need for 18 cycles of PCR already with 50ng of starting material.
We have used differing inputs for the same samples and we have not observed higher rates of clonality at MspI sites or a lower correlation (as highlighted under comment number one). We agree that lower PCR cycle numbers are advantageous, but the distribution of covered sites does not indicate an issue with over-amplification except for rare loci with extreme coverage results (see figure 5B in the manuscript). We have successfully prepared and sequenced ERRBS libraries with 15 cycles of PCR starting with 50 ng of DNA. We have included a notation in the protocol that advises that PCR cycle reduction is possible as further optimization of the protocol (protocol step 7). We recommend that novice users start with 18 cycles of PCR to ensure success with the protocol. Experienced users can reduce the number of PCR cycles and still achieve sufficient library product for sequencing. We have indicated that titration of the number of PCR cycles is possible in the paper for the readers' consideration. Finally, we have found that 18 cycles of PCR are necessary when starting with 5 ng of material.

6. Throughout the protocol, large reaction volumes and excessive amounts of enzymes are used, which makes the protocol unnecessarily expensive.
We appreciate this comment. The protocol submitted is the current approach used. The cost of the reagents and consumables used for a single ERRBS prep is $120.00. Compared to other next generation sequencing protocol kits used, such as RNAseq and exome capture, we do not find that the cost of the library preparation is excessive. The protocol is continuously undergoing improvements and this is an aspect that is assessed on an ongoing basis.

Reviewer #4:
The manuscript describes improved RRBS procedure for genome scale DNA methylation analysis.

Whole genome bisulfite sequencing is the ultimate method for studying DNA methylation genome wide. RRBS focuses on most relevant genomic regions at an affordable cost. ERRBS presented here is very similar to RRBS but with modifications which improve reproduciblity and coverage of additional relevant genomic regions.

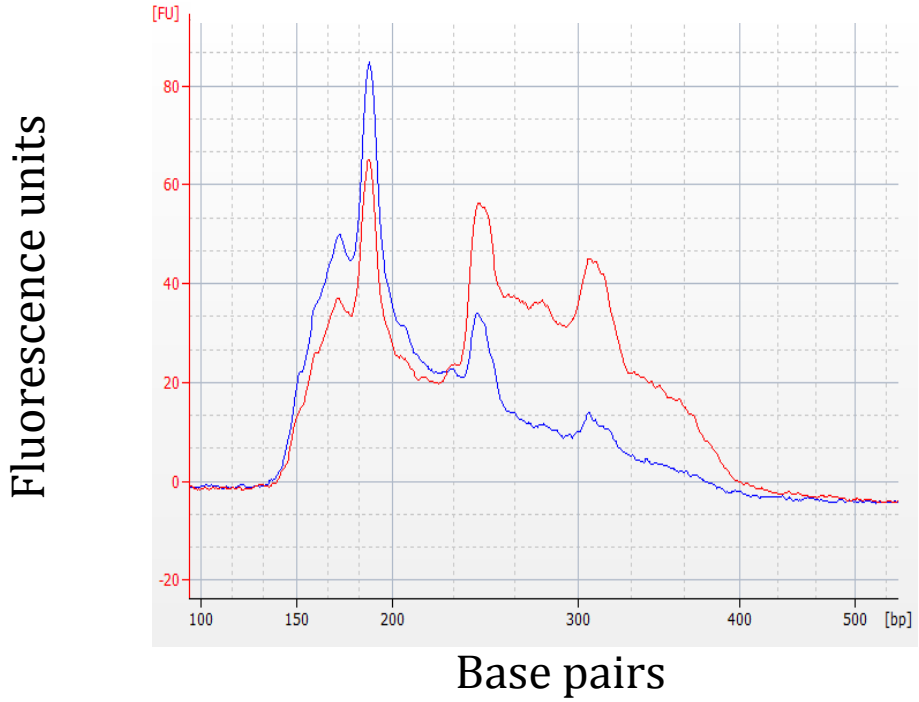
The protocol is described in detail in a user friendly way, paying attention to key considerations. Having a detailed protocol with video demonstrations and a complementary analysis guide is very useful for someone trying to setup this assay.

With this objective in mind I list some minor comments and some cases that additional orientation and tips can prove useful when setting up the protocol.
We thank reviewer four for his/her thorough input. Below are comments to address each point raised.

comments to authors
1. The protocol is described for 50ng DNA. If the authors have experimented with higher (or lower) amounts of input is will be helpful to provide some guidelines as to how this affects on the protocol downstream (for example on adaptor concentrations, size selection step, number of PCR cycles).
We thank the reviewer for this recommendation. We have revised the manuscript to include Table 3, which details how we recommend changing the protocol for various starting material quantities. We agree that this information will be useful to readers

2. The protocol is described for a single sample (which after size separation becomes two samples which are processed separately). It is common practice to pool samples together after adaptor ligation for subsequent processing. Instead, the authors choose to split each fraction into 4 amplification tubes. It will be interesting to know what led to this decision. In this case, I assume that the two size ranges have to be processed separately at least until amplification. But in theory, when several DNA samples are studied, it may be possible to pool samples together after size selection so that the high fragment size fraction from several samples are bisulfite converted and amplified together. This should reduce costs and reduce heterogeneity between samples introduced by processing. Assuming the authors considered this possibility, it will be helpful if they can share their conclusions, and if this possibility is practical, offer some guidelines (or full details) how can this be done.
We thank the reviewer for this comment. The rationale for the division of the two fragments is for the optimization of the downstream PCR amplification step. We have found that maintaining independent higher and lower fractions results in higher efficiency of PCR amplification and more equal representation of each in

the final library pool (please see figure below: red trace represents a pooled ERRBS library for sequencing prepared per our protocol; blue trace represents a library prepared as a single fraction isolated from size selection).



We agree that pooling lower and higher fragments from several samples together for sequencing can be cost effective, however only if limited coverage per CpG is a desired result. Furthermore, any pooling approach after size selection will be limited by the optimal input DNA quantity that the bisulfite conversion method used can utilize. The protocol submitted for publication is our current version in use. It is continuously under development and this recommended consideration for change in the protocol can be taken into consideration for further improvement. For the purposes of performing analysis to determine epigenetic polymorphisms as described by Landan et al.[2], it is necessary to obtain high coverage data and we recommend the use of the equivalent of single lanes per sample for such data acquisition.

The protocol describes pooling of the two fractions (high and low size) for one sample but does not address pooling of different libraries (different DNA sources) for running on one sequencing lane.

We thank the reviewer for this comment. While more cost effective, multiplexing several samples per lane will significantly reduce the number of CpGs covered and the depth of coverage for each CpG represented in the data. We have revised the manuscript to include additional information on a multiplexing approach that can be used should that be a desired approach (protocol step 12), as well as information on data yields possible from such an approach compared

to sequencing a single ERRBS library per lane (Figure 5 and Table 4).

3. The protocol emphasizes to importance of sequencing control unconverted sample (such as RNAseq ChIPseq) in parallel to bisulfite to ensure correct base calling. The term control lane suggests it is a different lane. The authors should clarify whether they mean that a control sample can be multiplexed with the bisulphite converted sample/s and run in the same lane.

We thank the reviewer for this comment and agree that further clarification could be useful to the readers. We have added additional details in protocol step 10 to address this topic, including the clarification that we are recommending the use of an independent lane for a control sample.

I could not find here a comment on the desired number of reads for the human/mouse genome, or reference to more elaborate examination of this issue.

We thank the reviewer for his/her comment and agree that such information would be helpful to the readers at large. We have included a figure with simulated estimates of CpGs covered per read counts for human genomes compared to sequencing a full lane per sample to address this question. Furthermore, we included data from our experience with multiplexed sequencing for the readers' reference (Figure 5 and Table 4).

Figure 3: images for good and bad preparations are presented. It will be useful to provide more guidance as to what to pay attention when assessing the outcome.

We thank the reviewer for this comment. We have added figure 3G to further visualize high quality library products appropriate for sequencing. We expanded the discussion to address key features of library quality control and potential failure criteria to consider. Furthermore, in the results, we have included assessment data from ERRBS sequencing runs (bisulfite conversion rates, cluster densities, pass filter rates, number of uniquely aligned reads per lane, alignment efficiency, the number of CpGs covered per sample and the average coverage at each CpG covered). Should there be any additional information the reviewer deems appropriate for inclusion, we would be happy to do so.

Minor Concerns:

1. Specify the PH when buffers are mentioned (for example item 1.4.1.1 - line 164)

We have included this edit in our updated manuscript

2. specify temperature for centrifugation (for example item 1.4.1.3 line 170)

We have included this edit in our updated manuscript

3. specify temperature for other steps (for example ethanol precipitation

step 1.4.2.1line 177)
We have included these edits in our updated manuscript

4. When using thermal cycler for incubations specify whether heat led should be on or off (lines 226,242,265)
We have included this edit in our updated manuscript

5. In item 4.3 (line 267) and elsewhere when magnetic beads (such as Agencourt AMPure XP) is used. the buffer concentration/volume affects size selection. It is therefore important to specify the volume/concentration of buffer used.
We have included this edit in our updated manuscript

1 Yu, M. *et al.* Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc*. **7** (12), 2159-70, doi: 10.1038/nprot.2012.137 (2012).

2 Landan, G. *et al.* Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet*. **44** (11), 1207-14, doi: 10.1038/ng.2442 (2012).

3 Boyle, P. *et al.* Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol*. **13** (10), R92, doi: 10.1186/gb-2012-13-10-r92 (2012).

4 Chatterjee, A., Rodger, E.J., Stockwell, P.A., Weeks, R.J., Morison, I.M. Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. *Journal of biomedicine & biotechnology*. **2012**, 741542, doi: 10.1155/2012/741542 (2012).

5 Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc*. **6** (4), 468 - 481, doi: 10.1038/nprot.2010.190 (2011).

Supplemental_Document_Multiplexing_protocol

Supplemental code file 1

Click here to download Supplemental code file (if applicable): Supplemental_code_file_1_JoVE_08252014.docx

Supplemental code file 2

Click here to download Supplemental code file (if applicable): Supplemental_code_file_2_JoVE_08252014.docx