Bisulfite oligonucleotide-capture sequencing for targeted base- and strand-specific absolute 5-methylcytosine quantitation

Dustin R. Masser · David R. Stanford · Niran Hadad · Cory B. Giles · Jonathan D. Wren · William E. Sonntag · Arlan Richardson · Willard M. Freeman

Received: 1 April 2016 / Accepted: 6 April 2016 / Published online: 18 April 2016
© American Aging Association 2016

Abstract Epigenetic regulation through DNA methylation (5mC) plays an important role in development, aging, and a variety of diseases. Genome-wide studies

of base- and strand-specific 5mC are limited by the extensive sequencing required. Targeting bisulfite sequencing to specific genomic regions through sequence

Electronic supplementary material The online version of this article (doi:10.1007/s11357-016-9914-1) contains supplementary material, which is available to authorized users.

D. R. Masser · D. R. Stanford · W. E. Sonntag ·
A. Richardson · W. M. Freeman
Department of Physiology, University of Oklahoma Health
Sciences Center, Oklahoma City, OK, USA

D. R. Masser
e-mail: dmasser@ouhsc.edu

D. R. Stanford
e-mail: dstanfor@ouhsc.edu

W. E. Sonntag
e-mail: William-Sonntag@ouhsc.edu

A. Richardson
e-mail: Arlan-richardson@ouhsc.edu

D. R. Masser · D. R. Stanford · N. Hadad · J. D. Wren ·
W. E. Sonntag · A. Richardson · W. M. Freeman
Reynolds Oklahoma Center on Aging, Oklahoma City, OK, USA

N. Hadad
e-mail: niran-hadad@ouhsc.edu

J. D. Wren
e-mail: jonathan-wren@omrf.org

D. R. Masser · D. R. Stanford · W. M. Freeman
Harold Hamm Diabetes Center, Oklahoma City, OK, USA

D. R. Stanford · J. D. Wren · W. E. Sonntag ·
A. Richardson · W. M. Freeman
Oklahoma Nathan Shock Center of Excellence in Basic Biology of
Aging, Oklahoma City, OK, USA

N. Hadad
Oklahoma Center for Neuroscience, Oklahoma City, OK, USA

C. B. Giles · J. D. Wren
Arthritis and Clinical Immunology Program, Oklahoma Medical
Research Foundation, Oklahoma City, OK, USA

C. B. Giles
e-mail: cory-giles@omrf.org

J. D. Wren
Department of Biochemistry and Molecular Biology, University of
Oklahoma Health Sciences Center, Oklahoma City, OK, USA

A. Richardson
Oklahoma City VA Medical Center, Oklahoma City, OK, USA

W. E. Sonntag · A. Richardson · W. M. Freeman
Department of Geriatric Medicine, University of Oklahoma Health
Sciences Center, Oklahoma City, OK, USA

W. M. Freeman (✉)
Department of Physiology OUHSC, BSMB 653, P.O. Box 26901,
Oklahoma City, OK 73126, USA
e-mail: wfreeman@ouhsc.edu

capture with complimentary oligonucleotide probes retains the advantages of bisulfite sequencing while focusing sequencing reads on regions of interest, enables analysis of more samples by decreasing the amount of sequence required per sample, and provides base- and strand-specific absolute quantitation of CG and non-CG methylation levels. As an example, an oligonucleotide capture set to interrogate 5mC levels in all rat RefSeq gene promoter regions (18,814) and CG islands, shores, and shelves (18,411) was generated. Validation using whole-genome methylation standards and biological samples demonstrates enrichment of the targeted regions and accurate base-specific quantitation of CG and non-CG methylation for both forward and reverse genomic strands. A total of 170 Mb of the rat genome is covered including 6.6 million CGs and over 67 million non-CG sites, while reducing the amount of sequencing required by ~85 % as compared to existing whole-genome sequencing methods. This oligonucleotide capture targeting approach and quantitative validation workflow can also be applied to any genome of interest.

Keywords Bisulfite sequencing · Rat · DNA methylation

Introduction

DNA methylation, specifically modified cytosine nucleotide bases with a methyl group on the 5th base position (5-methylcytosine or 5mC), is a central regulator of genome structure and gene expression. DNA methylation can be both a permanent and dynamic regulator of chromatin accessibility, and alterations in genomic DNA methylation patterns regulate changes in gene expression with development, aging, and in disease states (Jaenisch and Bird 2003; Jones 2012). Historically, 5mC was analyzed by pyrosequencing or affinity enrichment methods, which while useful, had significant limitations in quantitation, base- and strand-specificity, and genomic coverage. Only recently has technology and methodology matured to a point where DNA methylation can be assessed in a quantitatively rigorous manner across the whole genome. Examples include the development of deep sequencing (Lister et al. 2009), and high-throughput, low-input, bisulfite conversion protocols (Izzi et al. 2014). Additionally, bisulfite sequencing methods have advanced methylation analysis to absolute quantitation with base- and

strand-specific accuracy from the methods that generated relative quantitation of methylation over genomic regions (methylated DNA immunoprecipitation (meDIP), methyl-CpG binding domain (MBD)). While whole-genome bisulfite sequencing (WGBS) is a powerful tool for biomedical studies, the amount of data required for whole-genome studies that are quantitatively accurate and have the base-specific required coverage depth (~3Gb genome at the 20× (10× per strand)) (Ziller et al. 2015) is at least 150 Gb per sample, when accounting for losses during trimming and alignment. Therefore, performing adequately powered WGBS studies with multiple independent samples per treatment group is still generally unfeasible despite continued increases in sequencing outputs. A significant amount of WGBS data generated is uninformative since it comes from repeat regions and cannot be confidently aligned. Base-specific methods that target specific genomic loci, including bisulfite amplicon sequencing (BSAS) (Masser et al. 2013, 2015), pyrosequencing (Dupont et al. 2004), and MassArray (Ehrich et al. 2005) are more feasible but are limited in the amount of the genome that is covered (<10 kb) (Laird 2010). Techniques such as reduced representation bisulfite sequencing (RRBS) are commonly used; however, because RRBS coverage is dependent on the distribution of restriction enzyme sites, many regions of interest are not covered (~20–25 % of gene promoters and 50–70 % of CG island shores in the human and mouse genomes (Meissner et al. 2005)). Furthermore, restriction digestion may be inhibited in tissues such as the CNS with high non-CG methylation (e.g., *MspI* will not cut C^{me}CGG). Thus, while RRBS is a powerful discovery tool, comprehensive coverage of regions of interest and consistent methylation quantitation across multiple samples and multiple experimental groups can be difficult to achieve. Microarrays which combine bisulfite conversion with an existing SNP platform technology (Illumina 450K Array (Sandoval et al. 2011)) have also been widely used but cover relatively few CG and non-CG sites of the genome and are currently only available for human studies. Thus, while these existing techniques are useful, they do not combine genome-wide scope with sufficient sample throughput for genome-wide discovery and hypothesis-generating studies.

A potential solution to this technological gap is genome-wide methods that use sequence capture approaches in conjunction with bisulfite sequencing. Analogous to exome sequencing, this approach would

provide a good balance of genome coverage, necessary sequencing depth, absolute quantitation, and base- and strand-specific data. Methods which combine the power of bisulfite conversion with targeting to specific genomic locations have been developed, including oligonucleotide capture of target regions for bisulfite sequencing (Ivanov et al. 2013; Li et al. 2015). However, comprehensive capture sets of all promoters and CG islands in a genome and definitive demonstrations of capture enrichment, and most importantly, validation of quantitative accuracy have not been reported. Furthermore, capture probe sets are not commercially available for species other than humans. The demonstration of methylation state-independent capture and quantitative accuracy are critical as biasing of 5mC quantitation during capture enrichment is an obvious concern. Much like exome-capture sequencing, bisulfite oligonucleotide capture sequencing (BOCS), previously referred to as post-bisulfite conversion-targeted capture sequencing (Li et al. 2015), reduces the amount of sequencing data required while still interrogating relevant portions of the genome making it a cost-effective approach for genome-wide DNA methylation studies. BOCS entails generating a whole-genome bisulfite converted library followed by hybridization to sequence-specific probes, which capture targeted regions independent of methylation state. Probe sets can be designed against any genomic regions of interest (within limitations of having unique sequences to design against) for any species. This does entail synthesis and testing of the probe sets, as presented here.

In biomedical research, animal models are necessary to elucidate disease mechanisms and address efficacy and action of potential therapeutics. *Rattus norvegicus*, or the Norway rat, is a commonly used animal model that has important advantages in the fields of neuroscience, diabetes, and aging research, among others. Rat models are widely accepted as closer to human conditions, compared to mouse models, because of similar physiology and more complex behavioral endpoints (Iannaccone and Jacob 2009). Additionally, recent advances in transgenic technology make one of the prime advantages of mouse models, genetic manipulation, more readily available in the rat (Flister et al. 2015). Despite the utility of this species, tools to specifically analyze DNA methylation of the rat genome have lagged behind methods for human studies. Therefore, the rat

genome was used as an example of the BOCS technology for analyzing genome-wide, base-specific DNA methylation and as a demonstration of the steps to validate a new capture probe set. This set of capture probes was designed to assay each of the Rnor_6.0 RefSeq promoters (4000 bp upstream of transcription start sites (TSS) and 1000 bp downstream of the TSS) as well as all CG islands, shores, and shelves not present in repeat regions, across all autosomes and sex chromosomes. Whole-genome rat methylation standards of low (<5 %), medium (1:1 low/high), and high (>85 %) methylation were used to validate methylation-independent enrichment of the targeted regions and the base- and strand-specific quantitative accuracy of BOCS for the rat genome, a critical component of quantitative methods, which has not been previously reported. This is, to our knowledge, the first rat genome-wide oligo-capture approach for comprehensive analysis of DNA methylation in all gene promoters and CG islands and the first quantitative validation of the BOCS approach.

Results

Capture set design

As described in detail in the “Methods” section, probes were designed against annotated genic regions (4000 bp upstream and 1000 bp downstream of each RefSeq and non-overlapping Mammalian Gene Collection gene transcription start site) and CG island/shore/shelf (CGI) in the Rnor_6.0 genome build with CGI in repeat regions masked (Fig. 1a). Sequence capture probes were designed against both strands, independent of methylation state. In total, 4.2 million probes were synthesized for this design. Libraries were generated with methylated adapter sequences, bisulfite converted, hybridized, and captured with the oligonucleotide probes, followed by amplification of the captured library (Fig. 1b). In total, the design covers 168.8 Mb of the rat genome including multiple genic (promoters, exons, introns, and intergenic CGI regions) and CGI regions (island, shore, shelf) (Fig. 1c).

Whole-genome standard validation

The accuracy of the capture set was tested with whole-genome rat methylation standards with differing levels of methylation. These standards were first confirmed by pyrosequencing of the rat LINE-1 element with resulting values of 2 % for low, 52 % for medium, and 98 % for high. While suggesting accurate generation of the standards, this does not address CG (and CH) methylation across the genome. The genome-wide accuracy of the methylation standards was validated by WGBS. The standards were bisulfite converted and adapters were ligated prior to low cycle number PCR and library QC including sizing and quantity determination (Fig. 2a). Libraries were sequenced on the Illumina MiSeq to generate reads for whole-genome alignment. Whole-genome methylation levels were calculated based on the average methylation level across all CG sites from Bismark alignment to the Rnor_6.0 genome (Fig. 2b). Each whole-genome methylation level was quantified as expected with the low standard quantified as 2.2 %, medium as 58.1 %, and the high as 95.1 % CG methylation. These standards have minimal to no levels of non-CG methylation (≤ 1 %), and these sites can be used to estimate bisulfite conversion efficiency (≥ 99 %). In total, 5.25 million reads and 2 Gb of sequencing data (<1.5 million CG methylation calls) were generated per standard, yielding slightly less than $1\times$ average sequencing depth per base across the whole rat genome. Because the average quantitation of 5mC across the entire genome was all that was required to validate the standard, this approach provides an efficient but accurate method for analyzing whole genome methylation levels and validating these methylation standards.

BOCS performance

BOCS libraries were generated using methylation standards of low, medium, and high 5mC (as described above) and sequenced on the Illumina HiSeq 2500 in a paired-end 150 cycle fashion. For the three standards, the mean number of raw sequencing reads produced was 132.5 ± 10.2 M reads, corresponding to an average of 26 Gb per standard. Reads were trimmed for quality ($\geq Q23$ for an average Q score of 36 across all reads) prior to mapping to the whole Rnor_6.0 genome using Bismark (Krueger and Andrews 2011). After trimming, 86 % of the total reads, an average of 113.91 ± 8.36 M reads per sample, were retained and ~ 65 % of the

trimmed reads (74.13 ± 6.51 M reads per sample) uniquely mapped for each standard. Estimated duplicate reads from raw sequences across each library were 13.1 ± 0.9 %; however, after trimming and mapping, 0 % of reads mapped were reported as duplicates (Supplemental Table 1).

Targeted regions of the genome were then examined for coverage and 88 % of the targeted bases were found to have ≥ 1 sequencing depth. A small portion (12 %) of the probe regions had no or very low read depth across all of the standards. These regions were eliminated as failed probes. No immediately discernible pattern of failed probe regions was evident with equivalent GC content in these failed probe regions (44.7 %) as in good regions (45.4 %), and no sequence motifs were evident from failed regions. The efficiently captured and covered targeted regions (88 % of original design) maintained the targeted design distribution; introns 33.1 %, exons 7.2 %, promoters 39.0 %, and intergenic CGI regions 20.7 % (Fig. 3a). For CGI context, the distribution of efficiently capture regions was also similar to the targeted design distribution; CG island 6.3 %, shelf 38.2 %, shore 39.0 %, and non-CGI genic 16.2 % (Fig. 3a). The distribution of reads mapped to each of the targeted regions, by genic and CGI context, is presented in Supplemental Fig. 1 and demonstrates similar region representation to the target coverage in Fig. 3a.

Of the mapped reads (74.1 ± 6.5 M reads per sample), an average of 38 % were on target (28.3 ± 6.2 M reads per sample), representing a sixfold average enrichment of targeted regions from the whole genome (Supplemental Table 2). On-target reads are defined here by reads that map to a single genomic location that overlaps with targeted regions. The majority of off-target reads (82 %) randomly aligned to masked genomic regions (i.e., repeat elements) and the off-target regions had a similar GC content (42 %) as on-target regions (45 %). Sequence motif analysis for the off-target reads from masked regions and non-masked regions revealed no over-represented motifs (Supplemental Fig. 2).

Mean and cumulative frequency distributions of CG base coverage for on-target reads (corresponding to the “good” probes, 88 % of total probes (Supplemental file 1)), and off-target reads, independent of strand, were down selected to the same number of mapped reads per standard (66.3 ± 0.02 M reads) and averaged across the standards (Fig. 3b, c). Mean frequency and cumulative frequency distributions for each sample individually is presented in Supplemental Fig 3. Ninety-five percent

Fig. 1 Overview of BOCS method and target regions. **a** Schematic of targeted regions. All masked CG island regions were targeted which included the CG island and flanking shores (± 2 kb from CGI) and shelves (± 4 kb from CGI (± 2 kb from shores)). Targeted RefSeq promoter regions included the promoter (-4 kb from transcription start site (TSS)) and the initial intragenic

region ($+1$ kb from the TSS). **b** Bisulfite oligonucleotide capture sequencing workflow. Sequencing libraries are generated from genomic DNA and bisulfite converted. Converted libraries are hybridized to capture probes, and captured libraries are then amplified and sequenced. **c** Statistics for BOCS-targeted regions

(95.3 ± 0.4 %) of targeted bases (CGs) in the good target regions were covered by at least one read (Supplemental Table 3). This fraction of targeted regions covered does not diminish significantly until less than 15 M mapped reads are used (Supplemental Fig. 4), a level of coverage depth far below the suggested level for quantitative accuracy. Based on the cumulative and frequency distributions of coverage, 50 % of the targeted bases had a

mean read depth of 20x (Fig. 3c and Supplemental Table 2). Enrichment for the target regions is evident by the rightward shift between off-target and on-target base coverage (Fig. 3b, c).

As an assessment of sample variability in capture and enrichment, we processed 21 independent biological rat retinal genomic DNA samples through the BOCS protocol. After trimming, an average of 18.4 ± 2.4 M reads

Fig. 2 Validation of whole genome rat methylation standards. **a** Schematic of whole genome bisulfite sequencing workflow. Libraries are generated by bisulfite converting genomic DNA and ligating sequencing adaptors. Libraries are then amplified and sequenced. **b** Observed CG methylation from whole genome rat methylation standards (low, medium, and high) as measured by whole-genome bisulfite sequencing

Fig. 3 BOCS sequencing and capture metrics. **a** Capture region representation statistics for low, medium, and high methylation standards of BOCS data visualized as genic and CGI regions. Mean frequency (**b**) and cumulative distributions (**c**) for on-target (mean—black line, S.D.—blue shade) and off-target (mean—black dashed line, S.D.—gray shade) sequencing depth

across all targeted bases for low, medium, and high methylation standards. Mean frequency (**d**) and cumulative distributions (**e**) for on-target (mean—black line, S.D.—blue shade) and off-target (mean—black dashed line, S.D.—gray shade) sequencing depth across all targeted bases for 21 rat retinal samples; S.D. standard deviation

Fig. 4 BOCS quantitative validation **a** observed genome-wide CG methylation values from whole genome rat methylation standards (low, medium, and high) measured by BOCS. **b** Frequency of methylation sites, in millions (M), and bins of CG methylation values of each whole-genome standard (low, medium, and high) visualized for each strand (*top* or *bottom*). **c** Circos plot visualizing genome-wide methylation values from BOCS of (*outer circle* to *inner circle*) high, medium and low CG methylation represented on each autosome and in strand (*top* or *bottom*) orientation (scale 0–100 %). **d** Mean methylation percentages ($\pm$ S.D.) from each standard (low, medium, and high) across either all covered promoters (*top*) or all covered CG islands (*bottom*), binned into 10ths.

e Example regions (promoter, CG island, and a promoter with a CG island) with methylation level from each standard (low—*red*, medium—*green*, and high—*blue*). The hepatoma-derived growth factor-like 1 (Hdgfl1) promoter region is on chromosome 17 from bases 40,166,552–40,171,552. The CG island region is on chromosome 1 from bases 15,870,839–15,871,230 and contains 29 CG sites. The combined promoter and CG island region is from the Dpy30 (dpy-30 histone methyltransferase complex regulatory subunit) promoter located on chromosome 6 from bases 22,298,395–22,303,395. The CG island in this promoter contains 49 CG sites; S.D. standard deviation

were uniquely aligned (duplicates removed) per sample. Frequency and cumulative distributions of on-target reads demonstrate a reproducible enrichment (sixfold) for on-target regions, as noted by the rightward shift between off-target and of on-target distributions (Fig. 3d, e). Capture efficiency was highly reproducible across these samples as demonstrated by the minimal variation in on-target coverage. Each independent biological DNA sample mean frequency and cumulative frequency distributions are presented in Supplemental Fig. 3.

BOCS methylation quantitation

Genome-wide methylation levels of the BOCS libraries from the rat methylation standards were determined from average methylation levels across all CG sites (Fig. 4a). Cytosine methylation levels were reported from all cytosine contexts (CG, CHG, and CHH). The genome-wide CG methylation levels were near those of the expected values with the low standard quantified as 1.5 %, the medium standard as 51.5 %, and the high standard as 95.2 %. CHG and CHH genome-wide cytosine methylation were ~1 %, which was expected as the standards were methylated only at CG sites. The average bisulfite conversion efficiencies for each standard were 98.98 ± 0.04 % (Supplemental Table 4). Standard percent methylation values for each CG with a sequence depth of ≥ 10 were plotted on a frequency distribution (in millions [M]) (Fig. 4b). Accurate, strand-specific quantitation was demonstrated, and the expected quantitation of each methylation standard on each strand was correct with the low standard distribution having the highest frequency of sites between 0 and 10 % (1.8 ± 2.4 %), the medium standards having the highest distribution at 50 % (46.2 ± 16.1 %), and the high standard having the highest frequency of sites between 90 and 100 % (94.4 ± 11.5 %) CG methylation (Fig. 4b). Genome-wide CG methylation levels from each standard at sequencing depths of ≥ 10 were visualized (scale 0–100 % for each strand) on a circus (Krzywinski et al. 2009) plot showing correct mean quantitation for each standard regardless of the region or chromosome (Fig. 4c). Of note, the off-target reads maintained correct quantitation for each standard. This was evident when investigating off-target read mappings across the mitochondrial genome. A total of 415 CG sites on the mitochondrial genome were used for quantitation across each standard (CG sites with >10 sequencing depth). The levels of CG

methylation across these sites were 1.7, 43.4, and 92.4 % 5mC for low, medium, and high standards, respectively.

Finally, mean methylation levels of each standard across either promoters or CGIs (bins of 10) were plotted and showed accurate methylation quantitation (Fig. 4d). Of note is that in the CGIs, there was no shift in methylation quantitation between the islands and the shores and shelves, further demonstrating no bias in quantitation between regions with different GC contents. Furthermore, specific promoter (hepatoma-derived growth factor-like 1; *Hdgfl1*), CGI, and overlapping promoter and CGI (*Dpy30*) regions were shown to have correct methylation levels from each standard (Fig. 4e).

Discussion

This is the first bisulfite oligonucleotide capture sequencing (BOCS) probe set developed and validated to accurately quantify DNA methylation in a genome-wide and base- and strand-specific manner for all Rnor_6.0 RefSeq gene promoters and every masked CG island/shore/shelf. The presented approach to probe design and synthesis, validation with methylation standards, and bioinformatic analysis can be used for any genome of interest and for any selection of target regions. For example, while not currently annotated for rat, a probe set for enhancer regions could be designed and validated using the approach presented here. The BOCS approach yields genome-wide data and provides accurate, absolute quantitation in a base- and strand-specific manner for both CG sites as well as non-CG sites. This approach of genomic enrichment through in-solution hybridization greatly reduces the amount of sequencing required to obtain read depths sufficient for precise DNA methylation quantitation in a site- and strand-specific manner.

Capture efficiency

Validation of this method with whole-genome CG methylation standards demonstrates two important criteria: (1) enrichment of the targeted regions and (2) quantitative accuracy. As with any sequence capture approach, the goal is to focus sequencing on specific targeted regions without introducing any quantitative bias during capture. The majority of targeted regions (88 % of the

design) was covered and had sufficient sequencing depth for accurate methylation quantitation. The 12 % of targeted regions that captured poorly—regions with no or low numbers of mapped reads—did not contain any identifying sequence motifs that might cause failed capture, and the GC content was not different from the regions that captured efficiently. Probes demonstrating poor capture (Supplemental File 1) will be eliminated and re-designed in the future to increase the specificity of the hybridization step to targeted regions. This modification will increase the amount of on-target read mapping, effectively increasing the sequencing depth of target regions. Nonetheless, the capture efficiency was equivalent between samples with very little deviation (S.D. = 1.25 % in on-target reads) across all of the standards and the 21 biological samples. In this first iteration of BOCS for the rat genome, off-target reads observed can be attributed to the difficult nature of bisulfite sequencing analysis, where the genome is reduced to three bases which greatly decreases the amount of genomic complexity (Laird 2010). As quantitative accuracy of bisulfite sequencing is dependent on sequencing depth, having equivalent capture efficiency and therefore coverage depths aids in quantitation and allows for appropriate statistical tests to be carried out on the quantitative data. Future studies using this method should further optimize the hybridization step in order to maximize the enrichment of targeted regions and minimize off-target capture.

Quantitative accuracy

In order to quantitatively validate the BOCS capture probe set for the rat genome, whole genome CG methylation standards of low, medium, and high methylation were used. For each standard, the correct quantitation of 5mC was retained through capture demonstrating that no quantitative bias is introduced, a critical metric that has not been previously reported for targeted-capture studies (Allum et al. 2015; Ivanov et al. 2013; Li et al. 2015). For the low and high standards, quantitation differed between the WGBS, BOCS, and pyrosequencing data by <1 %, with the medium standard levels for the three methods differing slightly more (range = 6.6 %). Additionally, as demonstrated by the standards, there was consistent base-level quantitation that did not differ by genic or CGI context. Similarly to exome studies, there may be utility in the off-target captured regions, not contained in masked repeats, as well, which

maintain accurate quantitation. Analysis of DNA methylation can be carried out on regions, such as mitochondrial DNA, as has become common among researchers analyzing exome data (Griffin et al. 2014; Samuels et al. 2013). With the conflicting reports of mtDNA methylation, analysis of mtDNA methylation using BOCS can be unbiased (Bellizzi et al. 2013; Hong et al. 2013) since sequence is captured in a methylation-independent manner and reads that map off-target to the mitochondrial chromosome maintain correct quantitation.

Sequencing recommendations

This specific oligonucleotide capture set permits the analysis of 170 Mb of highly relevant genomic loci with theoretically only 1/16th of the sequencing required to perform whole-genome bisulfite sequencing. According to the data presented here, based on mapping efficiency and probe capture efficiency, to achieve 10× sequencing depth per strand for the target regions, approximately 16 Gb per sample is needed for BOCS, while whole-genome bisulfite sequencing requires ~125 Gb to cover those regions at the same depth. Practically, this means that BOCS only requires 1/8th of the sequencing required for whole-genome bisulfite sequencing given that capture is not completely efficient. This bisulfite oligonucleotide capture performs well in comparison to exome sequencing technologies, which target coding regions of the genome for analysis of variants (Chilamakuri et al. 2014).

Based on our validation and amount of sequencing performed, researchers using BOCS as a method for genome-wide methylation quantitation can adjust sequencing parameters accordingly based on the number of samples and the target depth of sequencing desired. This allows for some flexibility in the amount of data generated. For those investigations examining only CG methylation, sequencing depths of >10× (16 Gb per sample based on the results of this study) are sufficient to confidently (i.e., within ±10 %) quantify methylation (Masser et al. 2013). However, studies that require non-CG methylation quantitation as well will need to sequence at depths of >20× (32 Gb per strand, per sample) because these sites are often methylated at low levels (Ziller et al. 2015). Additionally, more sequencing and subsequently greater coverage depth are useful when greater precision in CG quantitation is desired (Masser et al. 2013). More sequencing, however, does not significantly increase the percent of targeted regions

covered (Supplemental Fig. 4). Additionally, probe sets can be made against other annotated genomic elements including enhancers and insulators, and non-coding RNA regions, or new sets can be designed for other species.

Validation strategy

The validation approach used here also offers a number of advantages over existing approaches. Use of whole-genome methylation standards is not routinely employed in bisulfite studies, but as demonstrated here, not only can whole-genome standards be generated but this allows for enrichment and quantitative accuracy of the method to be rigorously assessed as well. Most previous reports rely on comparisons to whole-genome bisulfite sequencing from the same sample. While useful, the ability to validate quantitation across the full range of methylation levels (from 0 to 100 %) can help identify any biases introduced and once standards are confirmed, resource-intensive whole-genome sequencing comparisons are not necessary. Additionally, adding the routine use of methylation standards to any methylation assay improves the analytical rigor and ultimately the reproducibility of epigenomic analyses.

Comparison to other technologies

As stated previously, an important rationale for BOCS is that the base- and strand-specific quantitative capability of bisulfite sequencing is retained while targeting sequencing to genomic regions of interest. BOCS has a number of advantages over the widely used affinity enrichment (e.g., MBD-Seq) and restriction enzyme (e.g., RRBS) methylation analysis methods designed to reduce sequencing (Altobelli et al. 2013; Bose et al. 2015; Hartung et al. 2012; Watson et al. 2015). Capture approaches such as MBD-Seq and MeDIP-Seq do not provide base-specific quantitation of 5mC, and quantitation is only relative between samples. Restriction digestion approaches such as RRBS methods can provide base-specific absolute quantitation but they are a sampling of the genome dependent on the locations of restriction sites. Thus, restriction digestion-based methods do not provide a comprehensive view of particular genomic elements of interest (i.e., only about 70 % of mouse gene promoters (Gu et al. 2011) are analyzed with RRBS) and are not tunable to specific genomic regions of interest. While these affinity

enrichment and restriction digestion methods have greatly advanced epigenomics, the ability of BOCS to select regions of interest (e.g., every protein coding gene promoter and all masked CGIs, as demonstrated here) and provide genome-wide, absolute base-specific methylation quantitation is a major advantage. BOCS provides another tool for epigenomic research that meets the needs of investigations seeking to examine specific regions of the genome and enables greater number of samples and coverage depths to be achieved. For pure discovery studies, RRBS and WGBS still have great utility. From a practical standpoint, BOCS is about 50 % more expensive (inclusive of library preparation and sequencing costs) than RRBS is due to the costs of the oligonucleotide probes. In comparison, WGBS is about 500 % more costly than BOCS is. The appropriate technology choice (e.g., BOCS, RRBS, or WGBS) for large-scale methylation sequencing studies will depend on the specifics of the study (amount of genome needed, number of samples, costs).

BOCS is able to not only generate data for CG methylation but also non-CG methylation and hemimethylation data because the strand-specific results of BOCS can be accurately mapped to non-CG and hemimethylation sites. Methods that fail to assay each strand cannot comprehensively quantify non-CG methylation, which is asymmetric and an emerging element of epigenetic regulation of gene expression (Kinde et al. 2015). Additionally, strand-specific RNA sequencing is becoming the gold standard for transcript expression analysis; therefore, integrating strand-specific data sets of DNA methylation and RNA expression has the potential to yield greater insights into transcript regulatory processes over conventional strand-independent methods. It should be noted that some previous sequence capture approaches have been reported that only capture one strand (Allum et al. 2015; Ivanov et al. 2013) but here, we have designed against both strands of the genome so that accurate strand-specific quantitation is achieved. This assay design complies with recommendations set by Krueger et al. for analysis of short bisulfite sequencing reads (Krueger et al. 2012). This includes the quality control for processing of short reads for bisulfite sequencing by proper bioinformatic tools (FastQC, CLC Genomics Workbench 8.0.1), proper alignment of trimmed reads (Bismark), and proper calling and processing of mapped reads (MethylKit) in order to consistently analyze bisulfite sequencing data. Therefore, in order to generate more consistent and

quantitatively precise genome-wide methylation data, BOCS represents a major advancement for epigenetic studies and should be preferentially used over the currently existing methods given its advantages.

Rat genome methylation analysis

The BOCS probe set described here is based on the most current Rnor_6.0 genome build, and targeting all promoter regions and masked CGIs allows for the most relevant and comprehensive analysis of DNA methylation in rat model systems while reducing the amount of sequencing required by ~85 % as compared to whole-genome bisulfite sequencing. The targeted regions in this probe set are not comprehensive for all regulatory regions (i.e., enhancer elements (not annotated in the rat genome) and long-noncoding RNAs) but are comprehensive for gene regulatory regions among promoters and CGIs. The target regions and capture probes can be expanded or reduced to meet experimental needs. The impact and application of BOCS in rat model systems for basic and translational biomedical research is of great interest. Using BOCS, genome-wide methylation levels can be quantified in an accurate and precise manner. Previously, quantifying genome-wide levels of DNA methylation in rats, or any other model system, aside from whole-genome bisulfite sequencing, has often been estimated by using genomic repeat regions (Asada et al. 2006). While this approach was feasible prior to deep sequencing, alterations in these repeat regions are affected by environmental factors (Martín-Núñez et al. 2014) and should not be over-interpreted as an accurate surrogate for whole-genome or genome-wide methylation levels. In addition to rigorous 5mC quantitation, the integration of BOCS methylation data with gene and transcript expression data from RNA-seq is of high relevance for investigating epigenetic mechanisms of gene expression regulation. Because BOCS comprehensively assays methylation in all annotated protein-coding genes, combining these datasets provides a systematic analysis of gene expression regulation as well as transcript expression regulation. The base-specificity allows for correlations on a base-by-base level to expression levels in addition to identifying differentially methylated regions that overlap with transcription factor binding sites.

In summary, these findings demonstrate the utility of the BOCS method for genome-wide, targeted DNA methylation analysis adding to the epigenetic strategies

available for biomedical researchers. Additionally the validation approach used can be a model for qualification of other capture probe sets developed in the future. An exciting future direction of BOCS methodology in general will be the application of this technology to other important animal models in biomedical research, such as non-human primates (King et al. 2012) and mice, which have been established as important model systems for disease research in the areas of neuroscience, substance abuse, and aging.

Methods

Probe design

The non-overlapping set of RefSeq (Pruitt et al. 2014) and Mammalian Gene Collection (Team MGCP 2009) genes (18,814) was downloaded from UCSC (Rosenbloom et al. 2015) (<http://hgdownload.soe.ucsc.edu/downloads.html>, ftp://hgdownload.soe.ucsc.edu/goldenPath/currentGenomes/Rattus_norvegicus/database/) for the most recent build of the rat genome (Rnor_6.0; <https://www.hgsc.bcm.edu/rat-genome-project>; July 2014 (Rosenbloom et al. 2015)). Transcription start sites (TSS) were used to set “promoter regions” at 4000 bp upstream plus 1000 bp downstream of each TSS. CG islands (Gardiner-Garden and Frommer 1987) not present in repeat regions (18,411), i.e., masked CG islands, were downloaded from UCSC (Rosenbloom et al. 2015), and CG island shores and shelves (Irizarry et al. 2009; Sandoval et al. 2011) were defined as 4000 bp up and downstream of the masked CG islands (Fig. 1a). This list of primary target regions was used for generation of capture probes. Tiled probes, ranging from 50 to 100 bp in length, were designed and synthesized by Roche Nimblegen against each strand and methylation state (Li et al. 2015). The resulting capture set comprises a merged set of 22,146 primary target regions, with 4.2 million strand-specific probes covering 168,839,413 bp (6.6 million CG sites and over 73 million Cs in total).

Qualification of standards

Commercially available rat whole-genome methylation standards of low and high CG methylation were generated by enzymatic treatment of whole-genome amplification and M. SssI treatment, respectively, of rat

genomic DNA (EpiGenDx). Medium standards were created through a 1:1 mixture of the high and low standards. These standards were used to generate whole-genome bisulfite sequencing libraries according to the manufacturer's instructions (EpiGnome, Illumina). Each standard (100 ng) was bisulfite converted using the EZ DNA Methylation-Lightning method (Zymo Research) and used for random-primed DNA synthesis and 5' tagging followed by 3'-end blocking and low cycle PCR to incorporate index sequences. After PCR amplification, libraries were purified using 1× AMPure XP beads (Beckman Coulter), sized by High Sensitivity DNA Chip (Agilent) and quantified by KAPA library quantification qPCR (KAPA Biosystems). Libraries were sequenced on the MiSeq (Illumina) at a final concentration of 12 pM using a MiSeq (Illumina) v3 reagent kit of paired-end 75 cycles. Bisulfite-converted reads were aligned to the Rnor_6.0 reference genome using Bismark with Bowtie2. Whole-genome methylation levels from the Bismark alignment were used as the validation of each standard's methylation level.

Bisulfite oligonucleotide capture sequencing validation

Rat whole-genome methylation standards of low (<5 %), medium (1:1 low/high), and high (>85 %) (EpiGenDx) CG methylations were used to generate whole-genome bisulfite-converted libraries (KAPA Biosystems) for enrichment using hybridization to oligonucleotide probes for regions of interest (Fig. 1b). Each DNA standard (1 µg) was sonicated to an average peak size of 200 bp using the Covaris M220 with the following settings: PIP(W) = 50, duty = 20 %, cycles/burst = 200, TT(s) = 175, vol(µl) = 50, and run at 20 °C (Covaris). Average fragment size was confirmed by capillary electrophoresis (DNA 1000, Agilent Bioanalyzer, Agilent). Libraries were generated through blunt end repair, A-tailing, and adapter ligation (KAPA Biosystems). After library clean-up and size selection, libraries were bisulfite converted using the EZ DNA Methylation-Lightning (Zymo Research) method and eluted in 20 µl of PCR-grade water. Bisulfite-converted libraries were then amplified (11 cycles) and purified using 1.8× AMPure XP beads (Beckman Coulter). Libraries were then hybridized to oligonucleotide probes for ~72 h at 47 °C in a thermocycler with a heated lid. Note that because of the long duration of the hybridization reaction, a 0.2-ml tube with water was

incubated for ~72 h in the same thermocycler at 47 °C to ensure little to no loss of volume over the hybridization reaction. After hybridization, captured libraries were washed and isolated using paramagnetic beads. After isolation, bisulfite-converted and captured libraries were amplified (15 cycles) and purified using 1.8× AMPure XP beads (Beckman Coulter) and eluted in 50 µl PCR-grade water. Size and concentrations of libraries were determined using a Bioanalyzer DNA 1000 chip and KAPA library quantification qPCR, respectively. Quantified libraries were sequenced on the HiSeq 2500 paired-end 150 cycle RapidRun (Illumina) at a final concentration of 12 pM.

Biological libraries (21 in total, one independent animal per library) were created from genomic DNA isolated (ZR-Duet, Zymo Research) from adult male Sprague-Dawley rat retinas. All animal experiments were conducted according to protocols approved by the University of Oklahoma Health Sciences Center Institutional Animal Care and Use Committee. Genomic DNA was quantified by fluorescent assays (Qubit, Invitrogen) and quality checked by capillary electrophoresis (Tape Station, Agilent). One microgram of gDNA was treated as above to create libraries and perform capture, as described above. Libraries were sequenced on the NextSeq (Illumina) in paired-end 150 cycle high-output run at a final concentration of 1.8 pM. Only capture efficiency data for these libraries is presented here, as quantitative comparisons of these data are part of a separate study.

Bioinformatics

Fastq files from each sequencing run were assessed for quality metrics with FastQC (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>) and then trimmed for quality with CLC Genomics Workbench 8.0.1. Trimming parameters were set at Q23, allow 4 Ns, trim 5' end by 3 nucleotides, trim 3' end by 4 nucleotides, remove Illumina universal adapter, and discard reads shorter than 32 nucleotides. This resulted in a total of 222,415,202 mapped reads with an average Q value of 36.8 and an average read length of 133 bp for the methylation standards. The 21 biological samples had a total of 442,193,159 mapped reads with an average Q value of 34.0 and an average read length of 117 bp. Trimmed reads were aligned and mapped with Bismark v0.13.0 (Krueger and Andrews 2011)/Bowtie2 version 2.2.5 (Langmead and Salzberg 2012), adjusting

alignment parameters for the highest sensitivity and against both strands. Individual mean frequency and cumulative frequency distribution plots were generated with Strand NGS 2.6 (Agilent). Data analysis and file management utilized NGSUtils (Breese and Liu 2013), samtools (Li et al. 2009), and bedtools (Quinlan and Hall 2010). MethylKit (Akalin et al. 2012), BS-Smooth (Hansen et al. 2012), ClustalX2 (Larkin et al. 2007), and WebLogo (Crooks et al. 2004) were also utilized for tertiary analysis.

Acknowledgments The authors wish to thank Dr. Graham Wiley, Dr. Patrick Gaffney, and the Oklahoma Medical Research Foundation Genomics facility for the sequencing services and John Ferguson for aiding in figure generation. This work was supported the Donald W. Reynolds Foundation, the National Institute on Aging [R01AG026607, P30AG050911], National Eye Institute [R01EY021716, T32EY023202], National Institute on Drug Abuse [R21DA029405], Oklahoma Center for Advancement of Science and Technology [HR14-174], and in part by an award from Harold Hamm Diabetes Center at the University of Oklahoma.

Author contributions D.R.M., D.R.S., and W.M.F. designed the study. D.R.M. generated the study data. D.R.M., D.R.S., N.H., C.B.G., J.D.W., and W.M.F. performed analysis of data. D.R.M., D.R.S., and W.M.F. wrote the manuscript. W.E.S. and A.R. provided intellectual input and design considerations. All authors reviewed and contributed to the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

- Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* 13:R87. doi:10.1186/gb-2012-13-10-r87
- Allum F et al. (2015) Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. *Nat Commun* 6:7211. doi:10.1038/ncomms8211
- Altobelli G, Bogdarina IG, Stupka E, Clark AJ, Langley-Evans S (2013) Genome-wide methylation and gene expression changes in newborn rats following maternal protein restriction and reversal by folic acid. *PLoS One* 8:e82989. doi:10.1371/journal.pone.0082989
- Asada K et al. (2006) LINE-1 hypomethylation in a choline-deficiency-induced liver cancer in rats: dependence on feeding period. *J Biomed Biotechnol* 2006:17142. doi:10.1155/JBB/2006/17142
- Bellizzi D, D'Aquila P, Scafione T, Giordano M, Riso V, Riccio A, Passarino G (2013) The control region of mitochondrial DNA shows an unusual CpG and non-CpG methylation pattern. *DNA Res Int J Rapid Publ Rep Genes Genomes* 20:537–547. doi:10.1093/dnares/dst029
- Bose R et al. (2015) Tet3 mediates stable glucocorticoid-induced alterations in DNA methylation and Dnmt3a/Dkk1 expression in neural progenitors. *Cell Death Dis* 6:e1793. doi:10.1038/cddis.2015.159
- Breese MR, Liu Y (2013) NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics* 29:494–496. doi:10.1093/bioinformatics/bts731
- Chilamakuri CS et al. (2014) Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics* 15:449. doi:10.1186/1471-2164-15-449
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. doi:10.1101/gr.849004
- Dupont JM, Tost J, Jammes H, Gut IG (2004) De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal Biochem* 333:119–127. doi:10.1016/j.ab.2004.05.007
- Ehrich M et al. (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* 102:15785–15790. doi:10.1073/pnas.0507816102
- Flister MJ, Prokop JW, Lazar J, Shimoyama M, Dwinell M, Geurts A (2015) Guidelines for establishing genetically modified rat models for cardiovascular research. *J Cardiovasc Transl Res* 8:269–277. doi:10.1007/s12265-015-9626-4
- Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196:261–282
- Griffin HR et al. (2014) Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. *Genet Med* 16:962–971. doi:10.1038/gim.2014.66
- Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc* 6:468–481. doi:10.1038/nprot.2010.190
- Hansen KD, Langmead B, Irizarry RA (2012) BSsmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* 13:R83. doi:10.1186/gb-2012-13-10-r83
- Hartung T et al. (2012) Diametrically opposite methylome-transcriptome relationships in high- and low-CpG promoter genes in postmitotic neural rat tissue. *Epigenetics Off J DNA Methylation Soc* 7:421–428. doi:10.4161/epi.19565
- Hong EE, Okitsu CY, Smith AD, Hsieh CL (2013) Regionally specific and genome-wide analyses conclusively demonstrate the absence of CpG methylation in human mitochondrial. *DNA Mol Cell Biol* 33:2683–2690. doi:10.1128/MCB.00220-13
- Iannaccone PM, Jacob HJ (2009) Rats! *Dis Model Mech* 2:206–210. doi:10.1242/dmm.002733
- Irizarry RA et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41:178–186. doi:10.1038/ng.298

- Ivanov M, Kals M, Kacevska M, Metspalu A, Ingelman-Sundberg M, Milani L (2013) In-solution hybrid capture of bisulfite-converted DNA for targeted bisulfite sequencing of 174 ADME genes. *Nucleic Acids Res* 41:e72. doi:[10.1093/nar/gks1467](https://doi.org/10.1093/nar/gks1467)
- Izzi B, Binder AM, Michels KB (2014) Pyrosequencing evaluation of widely available bisulfite conversion methods: considerations for application. *Med Epigenetics* 2:28–36. doi:[10.1159/000358882](https://doi.org/10.1159/000358882)
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33(Suppl):245–254. doi:[10.1038/ng1089](https://doi.org/10.1038/ng1089)
- Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13:484–492. doi:[10.1038/nrg3230](https://doi.org/10.1038/nrg3230)
- Kinde B, Gabel HW, Gilbert CS, Griffith EC, Greenberg ME (2015) Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A* 112:6800–6806. doi:[10.1073/pnas.1411269112](https://doi.org/10.1073/pnas.1411269112)
- King GD, Rosene DL, Abraham CR (2012) Promoter methylation and age-related downregulation of Klotho in rhesus monkey. *Age (Dordr)* 34:1405–1419. doi:[10.1007/s11357-011-9315-4](https://doi.org/10.1007/s11357-011-9315-4)
- Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 27:1571–1572. doi:[10.1093/bioinformatics/btr167](https://doi.org/10.1093/bioinformatics/btr167)
- Krueger F, Kreck B, Franke A, Andrews SR (2012) DNA methylome analysis using short bisulfite sequencing data. *Nat Methods* 9:145–151. doi:[10.1038/nmeth.1828](https://doi.org/10.1038/nmeth.1828)
- Krzywinski M et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645. doi:[10.1101/gr.092759.109](https://doi.org/10.1101/gr.092759.109)
- Laird PW (2010) Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 11:191–203. doi:[10.1038/nrg2732](https://doi.org/10.1038/nrg2732)
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. doi:[10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923)
- Larkin MA et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. doi:[10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404)
- Li H et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. doi:[10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352)
- Li Q et al. (2015) Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. *Nucleic Acids Res* 43:e81. doi:[10.1093/nar/gkv244](https://doi.org/10.1093/nar/gkv244)
- Lister R et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322. doi:[10.1038/nature08514](https://doi.org/10.1038/nature08514)
- Martin-Nunez GM et al. (2014) Methylation levels of the SCD1 gene promoter and LINE-1 repeat region are associated with weight change: an intervention study. *Mol Nutr Food Res* 58: 1528–1536. doi:[10.1002/mnfr.201400079](https://doi.org/10.1002/mnfr.201400079)
- Masser DR, Berg AS, Freeman WM (2013) Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics Chromatin* 6:33. doi:[10.1186/1756-8935-6-33](https://doi.org/10.1186/1756-8935-6-33)
- Masser DR, Stanford DR, Freeman WM (2015) Targeted DNA methylation analysis by next-generation sequencing. *J Vis Exp JoVE*. doi:[10.3791/52488](https://doi.org/10.3791/52488)
- Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 33:5868–5877. doi:[10.1093/nar/gki901](https://doi.org/10.1093/nar/gki901)
- Pruitt KD et al. (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* 42:D756–D763. doi:[10.1093/nar/gkt1114](https://doi.org/10.1093/nar/gkt1114)
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. doi:[10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033)
- Rosenbloom KR et al. (2015) The UCSC genome browser database: 2015 update. *Nucleic Acids Res* 43:D670–D681. doi:[10.1093/nar/gku1177](https://doi.org/10.1093/nar/gku1177)
- Samuels DC, Han L, Li J, Quanguo S, Clark TA, Shyr Y, Guo Y (2013) Finding the lost treasures in exome sequencing data. *Trends Genet* 29:593–599. doi:[10.1016/j.tig.2013.07.006](https://doi.org/10.1016/j.tig.2013.07.006)
- Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics Off J DNA Methylation Soc* 6:692–702
- Team MGCP et al. (2009) The completion of the mammalian gene collection (MGC). *Genome Res* 19:2324–2333. doi:[10.1101/gr.095976.109](https://doi.org/10.1101/gr.095976.109)
- Watson CT, Szutorisz H, Garg P, Martin Q, Landry JA, Sharp AJ, Hurd YL (2015) Genome-wide DNA methylation profiling reveals epigenetic changes in the rat nucleus accumbens associated with cross-generational effects of adolescent THC exposure. *Neuropsychopharmacology*. doi:[10.1038/npp.2015.155](https://doi.org/10.1038/npp.2015.155)
- Ziller MJ, Hansen KD, Meissner A, Aryee MJ (2015) Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat Methods* 12:230–232. doi:[10.1038/nmeth.3152231](https://doi.org/10.1038/nmeth.3152231) p following 232