NGS Methyl-Sequencing

NEBNEXT® ENZYMATIC METHYL-SEQ (EM-SEQ™)

Heads up!

NEW ENGLAND BioLabs®

be **INSPIRED**
drive **DISCOVERY**
stay **GENUINE**

There's a new alternative to bisulfite sequencing – NEBNext® Enzymatic Methyl-seq (EM-seq™)

The identification of cytosine modifications within genomes, especially 5-methylcytosine (5-mC) and 5-hydroxymethylcytosine (5-hmC), is important as they are known to have an impact on gene expression. Generally, low levels of methylation near transcription start sites are associated with higher transcription levels, while genes with regulatory regions containing high levels of cytosine modification are expressed at lower levels.

The ability to analyze a complete methylome is important for studying diseases, including those associated with cancer, metabolic disorders and autoimmune diseases. Unfortunately, the current technologies for investigating 5-mC and 5-hmC are sub-optimal and do not permit a thorough evaluation of methylomes.

BISULFITE SEQUENCING

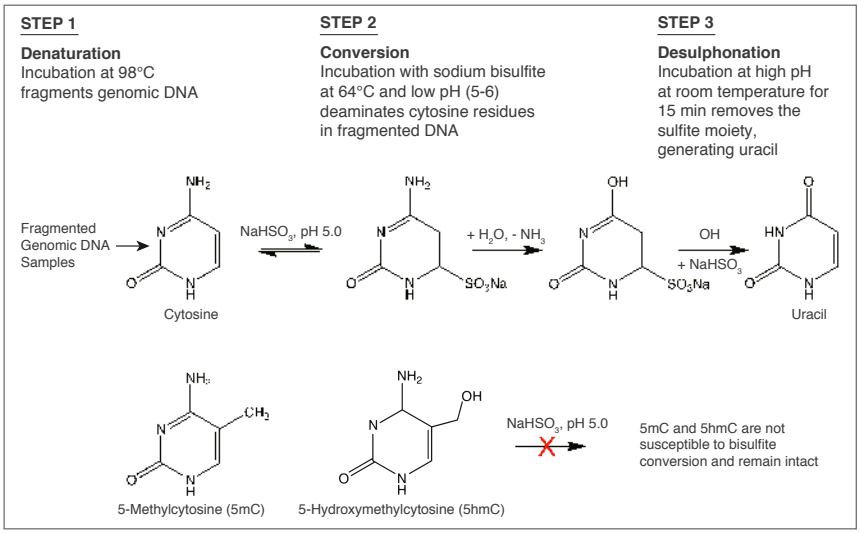
To date, the gold standard in methylome mapping has been bisulfite sequencing. In this method, DNA is chemically treated with sodium bisulfite, which results in the conversion of unmethylated cytosines to uracils, and the resulting uracils are ultimately sequenced as thymines. In contrast, the modified cytosines, 5-mC and 5-hmC, are resistant to bisulfite conversion, and are sequenced as cytosines (1).

While the preparation of bisulfite libraries is relatively straightforward, the libraries have uneven genome coverage and therefore suffer from incomplete representation of cytosine methylation across genomes. This uneven coverage is the result of DNA damage and fragmentation, which is caused by the extreme temperatures and pH during bisulfite conversion. Sequenced bisulfite libraries typically have skewed GC bias plots, with a general under-representation of G- and C-containing dinucleotides and over-representation of AA-, AT- and TA-containing dinucleotides, as compared to a non-converted genome (2).

Therefore, the damaged libraries do not adequately cover the genome, and can include many gaps with little or no coverage. Increasing the sequencing depth of these libraries can recover some missing information, but at steep sequencing costs. These bisulfite library limitations have driven the development of new approaches for studying methylomes.

Bisulfite conversion overview

Sodium bisulfite treatment of DNA converts cytosine to 5,6-dihydrocytosine-6-sulphonate, which is converted to 5,6-dihydrouracil-6-sulphonate, and then desulphonated to uracil. In contrast 5-mC and 5-hmC are not susceptible to bisulfite treatment and remain intact.



Advantages of EM-seq™

- Superior sensitivity of detection of 5-mC and 5-hmC
- Greater mapping efficiency
- More uniform GC coverage
- Detect more CpGs with fewer sequence reads
- Uniform dinucleotide distribution
- Larger library insert sizes
- High-efficiency library preparation
- Allows for lower input amounts with a range of inputs from 10–200 ng
- Conversion module also available separately

Customer Feedback:

> *(…) it enables us to determine in a precise and DNA sparing way the cytosine methylation status even at low integrity DNA. If bisulfite conversion were the only approach to apply, we would definitely fail to generate relevant results. The cool, biochemical approach (…) opens new avenues to explorations of methylation at intact long DNA fragments.*
>
> *Dr. Vladimir Benes, Head Genomics Core Facility at EMBL Heidelberg*

> *Enzymatic conversion of EM-Seq is THE alternative and our comparisons clearly showed, that the quality of data obtained is better than with conventional bisulfite conversion. We observed excellent conversion efficiency and extraordinary mapping rates.*
>
> *Dr. Alexander Vogt, Sequencing Specialist, Vienna BioCenter Core Facilities*

EM-seq WORKFLOW

In the EM-seq workflow, as with WGBS libraries, the first step is library construction from sheared DNA. For EM-seq, standard input amounts can range from 10-200 ng of sheared DNA, and a modified protocol is also available for input amounts as high as 500 ng. This is followed by two sets of enzymatic conversion steps to differentiate unmethylated cytosines from 5-mC and 5-hmC. Finally, libraries are PCR amplified before sequencing.

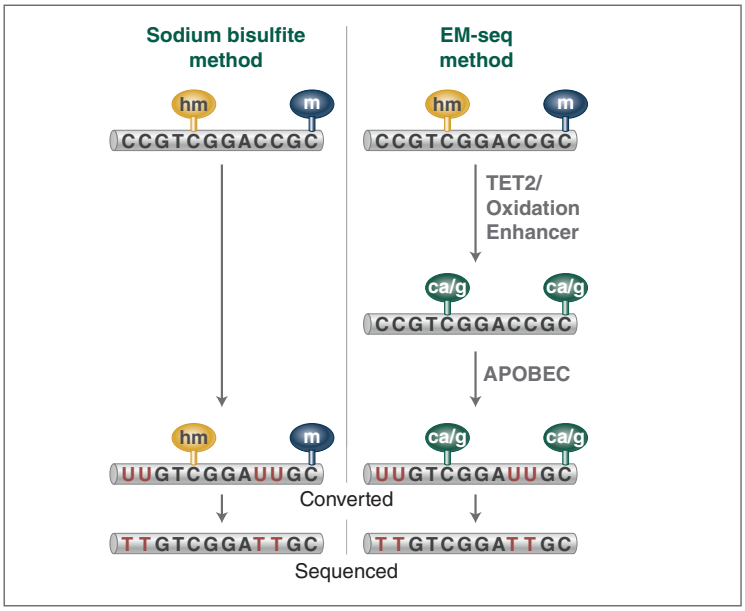
Sodium bisulfite treatment of DNA results in the deamination of cytosines to uracils, while the modified forms of cytosine (5-mC and 5-hmC) are not deaminated. When bisulfite treated DNA is PCR amplified, uracils are replaced by thymines, and 5-mC and 5-hmC are replaced by cytosines. Once sequenced, unmethylated cytosines are represented by thymines and 5-mC and 5-hmC are represented by cytosines. By comparing sequences to reference sequences (C/T and G/A converted genome), the methylation status can be assessed.

The first EM-seq conversion step uses TET2 and an Oxidation Enhancer to protect modified cytosines from downstream deamination. TET2 enzymatically oxidizes 5-mC and 5-hmC through a cascade reaction into 5-carboxycytosine [5-methylcytosine (5-mC) → 5-hydroxymethylcytosine (5-hmC) → 5-formylcytosine (5-fC) → 5-carboxycytosine (5-caC)]. This protects 5-mC and 5-hmC from deamination. 5-hmC can also be protected from deamination by glucosylation to form 5-ghmC using the Oxidation Enhancer.

The second enzymatic step uses APOBEC, which deaminates cytosine but does not affect 5-caC and 5-ghmC.

The resulting converted sequence is the same as that for bisulfite-treated DNA and so can be analyzed in the same way. Typical aligners used to analyze data include, but are not limited to, Bismark and bwa-meth.

EM-seq and sodium bisulfite conversion methods



NEBNext EM-seq Kit Workflow

EM-seq utilizes two enzymatic steps to differentiate between modified and unmodified cytosines.

Genomic DNA Shearing

Input is 10 –200 ng of genomic DNA, sheared to 300 bp

End Repair/dA-Tailing

DNA is end-repaired and dA-tailed

EM-seq Adaptor Ligation

DNA is ligated to the EM-seq adaptors

NEBNext® Ultra™ II reagents

Oxidation of 5mC and 5hmC

TET2 and Oxidation Enhancer protect 5mC/5hmC from deamination

Deamination of C to U

APOBEC deaminates cytosines to uracils; oxidized forms of 5mC/5hmC are not deaminated

PCR Amplification

Library amplification using NEBNext Q5U Master Mix and NEBNext index primers

Sequencing

Sequencing on the Illumina® platform

ENZYMATIC METHYL-SEQ – A NEW APPROACH

The enzymatic methyl-seq workflow developed at NEB provides a much-needed alternative to bisulfite sequencing. This method relies on the ability of APOBEC to deaminate cytosines to uracils. Unfortunately, APOBEC also deaminates 5-mC and 5-hmC, making it impossible to differentiate between cytosine and its modified forms (3,4).

In order to detect 5-mC and 5-hmC, this method also utilizes TET2 and an Oxidation Enhancer, which enzymatically modify 5-mC and 5-hmC to forms that are not substrates for APOBEC. The TET2 enzyme converts 5-mC to 5-caC and the Oxidation Enhancer converts 5-hmC to 5-ghmC (5,6,7). Ultimately, cytosines are sequenced as thymines and 5-mC and 5-hmC are sequenced as cytosines, thereby protecting the integrity of the original 5-mC and 5-hmC sequence information.

The NEBNext Enzymatic Methyl-seq Kit (EM-seq™) combines NEBNext® Ultra™ II reagents with these two enzymatic steps to construct Illumina® libraries that accurately represent 5-mC and 5-hmC within the genome. Converted libraries are amplified using NEBNext Q5U DNA polymerase. EM-seq libraries result in a more accurate representation of the methylome, with minimal DNA fragmentation or biases when compared to whole genome bisulfite sequencing (WGBS).

The combination of the Ultra II reagents for library prep and the EM-seq conversion allows for lower input amounts compared to most WGBS workflows, with a range of inputs from 10 – 200 ng.

EM-SEQ PERFORMANCE

Intact DNA

Several pieces of data suggest that the process of generating EM-seq libraries does not damage DNA in the same way as bisulfite sequencing. EM-seq libraries give higher PCR yields despite using fewer PCR cycles for all DNA input amounts, indicating that less DNA is lost during enzymatic treatment and library preparation, as compared to WGBS. Reduced PCR cycles, in turn, translates into more complex libraries and fewer PCR duplicates during sequencing (data not shown). EM-seq libraries also have larger insert sizes than WGBS, which further supports the fact that DNA remains intact.

EM-seq Libraries Have Reduced Bias

The preservation of DNA integrity is also demonstrated by the GC bias graphs, and the dinucleotide coverage distribution graph. Both of these figures indicate that reduced bias is associated with the EM-seq libraries. The EM-seq libraries have a flat GC bias distribution with even coverage at both GC and AT rich regions, and do not display a preference for any dinucleotide combination. This is in stark contrast to WGBS, which shows a skewed GC bias profile along with the previously mentioned dinucleotide biases. Reduced library bias improves the mapping and therefore coverage of CpGs.

CpG Detection

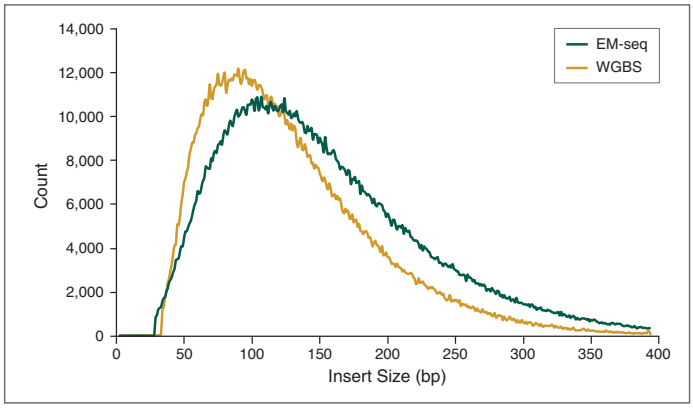
Human DNA is methylated almost exclusively in CpG contexts. EM-seq global CpG methylation levels for human NA12878 DNA are consistent with WGBS libraries, indicating that EM-seq libraries accurately detect methylation. The more striking difference between EM-seq and WGBS libraries becomes apparent when the focus is shifted to CpG coverage. EM-seq libraries detect more CpGs to a higher depth of coverage than WGBS libraries. The ability to detect more CpGs at a greater depth also increases confidence in the data and leads to more accurately defining methylation within a region of interest. This in turn aids in detecting methylation changes in diseased states such as cancer. In addition, increased CpG coverage has an economic impact – with more CpGs detected using the same number of reads compared to WGBS, EM-seq represents significant cost-savings.

Potential Applications

In addition to making Illumina libraries, there are other potential applications for the EM-seq technology. Many of these applications already exist, but can now be improved upon because of the intact nature of enzymatically-converted DNA and the accuracy of CpG detection. Lower input DNA is also a driving factor for some of these applications. Converted DNA can be detected on arrays, and can be used for target enrichment, reduced representation-type libraries or amplicon detection. Different types of DNA inputs, such as low input cell free DNA (cfDNA) or damaged FFPE DNA, can also be used.

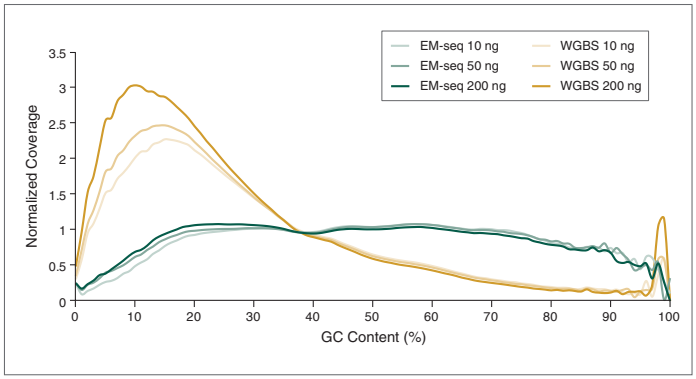
NEBNext Enzymatic Methyl-seq (EM-seq) libraries have larger inserts

EM-seq library insert sizes are larger than whole genome bisulfite sequencing (WGBS) libraries. Library insert sizes were determined using Picard 2.18.14. The larger insert size indicates that EM-seq does not damage DNA as bisulfite treatment does.



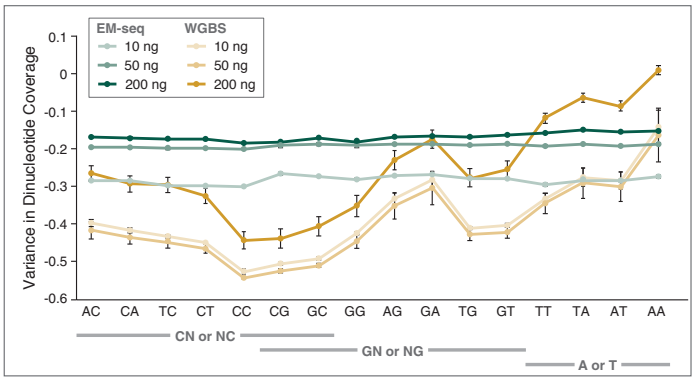
EM-seq has superior uniformity of GC coverage

GC coverage was analyzed using Picard 2.18.14 and the distribution of normalized coverage across different GC contents of the genome (0-100%) was plotted. EM-seq libraries have significantly more uniform GC coverage, and lack the AT over-representation and GC under-representation typical of WGBS libraries.



Dinucleotide coverage distribution

Dinucleotide coverage distribution for EM-seq and WGBS libraries showing the variance in coverage for dinucleotides in the reads when compared to unconverted Ultra II library dinucleotide distribution. EM-seq libraries show even coverage across all dinucleotide combinations compared to WGBS. C-containing dinucleotides are underrepresented in WGBS libraries and A/T containing dinucleotides are overrepresented.

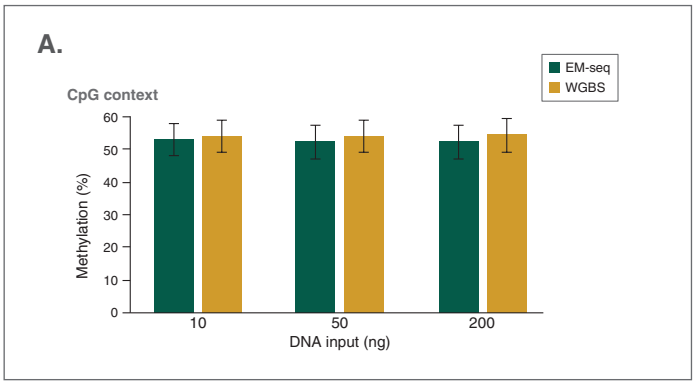


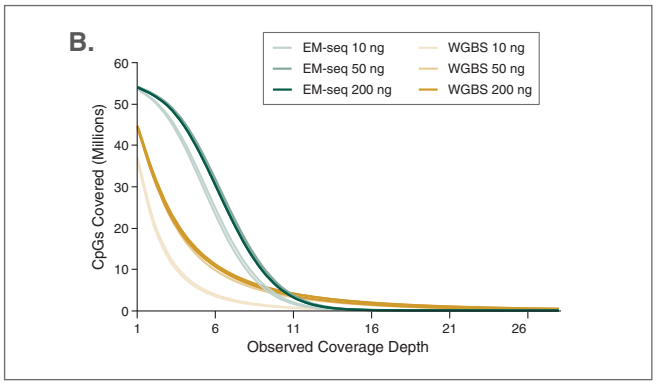
EM-seq identifies detect more CpGs to a higher depth of coverage than WGBS

10, 50 and 200 ng Human NA12878 genomic DNA was sheared to 300 bp using the Covaris S2 instrument and used as input into EM-seq and WGBS protocols. For WGBS, NEBNext Ultra II DNA was used for library construction, followed by the Zymo Research EZ DNA Methylation-Gold Kit for bisulfite conversion. Libraries were sequenced on an Illumina NovaSeq® 6000 (2 x 100 bases). 324 million paired end reads were aligned to hg38 using bwa-meth 0.2.2.

A: Methyl Dackel was used to determine methylation levels, which were found to be similar between EM-seq and WGBS.

B: Coverage of CpGs with EM-seq and WGBS libraries was analyzed, and each top and bottom strand CpGs were counted independently, yielding a maximum of 56 million possible CpG sites. EM-seq identifies more CpGs at lower depth of sequencing.





CONCLUSION

Bisulfite sequencing, while commonly used, is sub-optimal in detecting 5-mC and 5-hmC – large amounts of DNA are needed, DNA can be damaged, and sequences are biased towards AT-rich regions. Other methods that couple chemical or enzymatic treatment with bisulfite sequencing also share similar limitations. EM-seq provides the first commercially-available, non-bisulfite method that comprehensively addresses the limitations of bisulfite sequencing and represents a new opportunity for more complete methylome analysis. EM-seq libraries are not damaged and have longer inserts, higher PCR yields with fewer PCR cycles, and lack biases associated with GC content. More CpGs are identified with greater coverage depth using EM-seq, as compared to WGBS. These advantages all contribute to EM-seq having more usable sequencing data when comparing the same number of reads for EM-seq and WGBS, which ultimately reduces sequencing costs. EM-seq is the only commercially-available alternative to bisulfite sequencing that provides an effective method for accurate and comprehensive detection of 5-mC and 5-hmC across the genome, and offers a new, more accurate alternative for studying disease states.

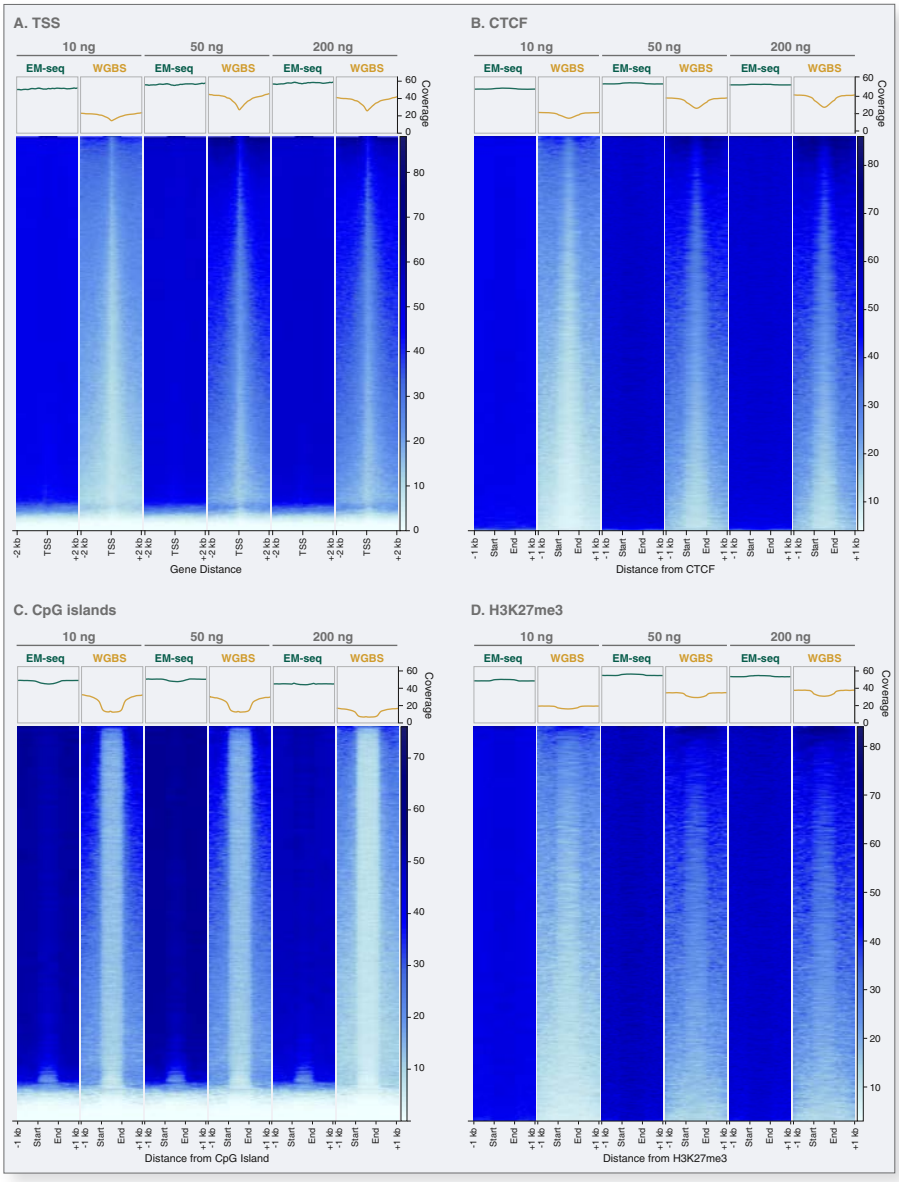
References
1. Harris R.A., et al. (2010) *Nat Biotechnol.* 28, 1097–1105.
2. Olova, N., et al. (2018) *Genome Biology,* 19: 33.
3. Carpenter, M.A., et al. (2012) *J. Biol.Chem.* 287, 34801–34808.
4. Wijesinghe, P. and Bhagwat, A.S. (2012) *Nucl. Acids Res.* 40, 9206–9217.
5. Josse, J and Kornberg, A. (1962) *J. Biol. Chem.* 237, 1968–1976.
6. Tomaschewski, J., et al. (1985) *Nucleic Acids Res.*13 (21): 7551–7568.
7. Schutsky, E.K., et al. (2017) *Nucleic Acids Res.* 45, 7655–7665.

ADDITIONAL DATA

TRANSCRIPTIONAL CONTROL FEATURES

In expressed genes, the regions at transcription start sites are generally unmethylated, while genes that are expressed at lower levels generally have high levels of cytosine methylation in regulatory regions. It is therefore important to be able to accurately define transcriptional control elements. However, since, as described above, WGBS data traditionally has lower C-containing dinucleotide representation, this results in the potential loss of CpG data around important transcription control elements. The distribution of coverage 1-2kb around several important transcriptional elements was examined, for EM-seq and WGBS: Transcription start sites (TSS) (A), CTCF transcriptional repressor protein binding site (B), CpG islands (C) and H3K27me3 histone methylation site (D). The more complete landscape, provided by EM-seq methodology, around these important sites, due to both the greater levels of coverage overall, and the greater uniformity of coverage, enable more confident analysis.

Examination of CpG coverage and methylation around transcription start sites is shown below. With EM-seq, coverage of CpGs is both significantly higher than WGBS in these 4kb windows, and also strikingly uniform, lacking the dip in coverage close to the transcription start site (TSS) itself characteristic of WGBS. The information on methylation garnered from EM-seq at 8X stringency also aligns more closely with expected methylation patterns, with a reduction in methylation levels close to the TSS.



Enhanced representation of transcriptional control features using EM-seq

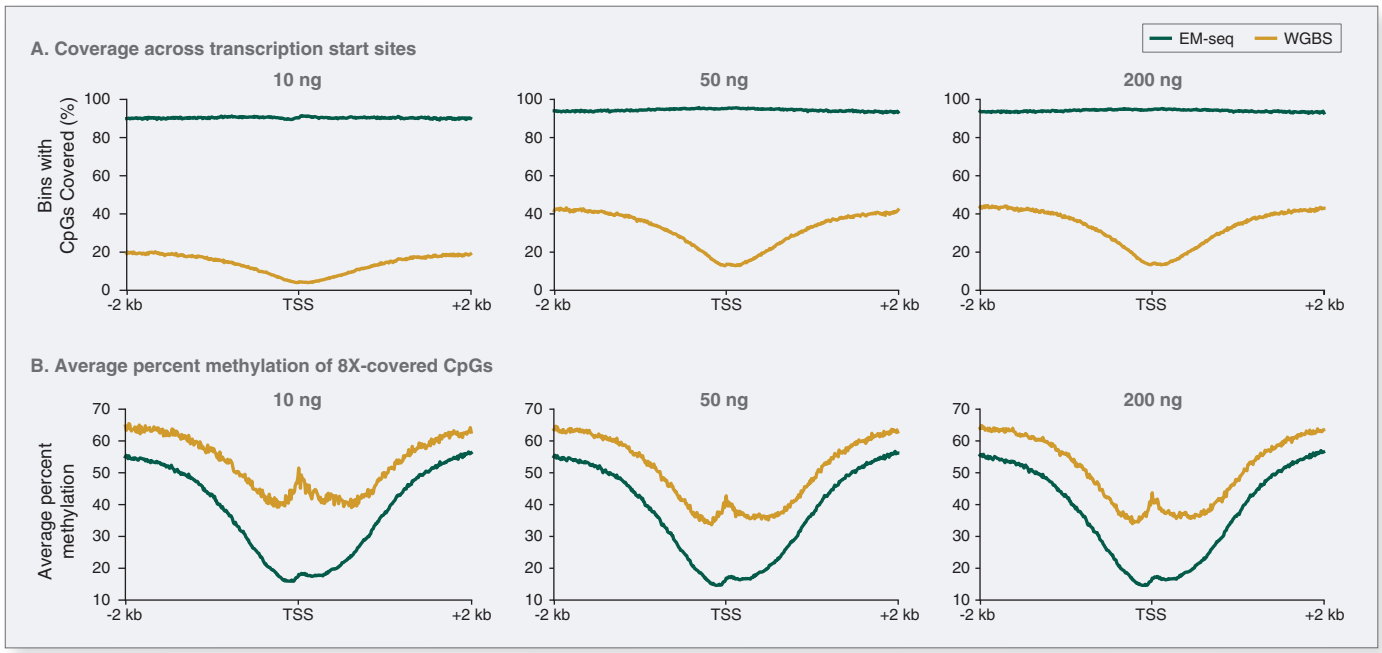
The lower C-containing dinucleotide representation associated with WGBS results in the potential loss of methylated and non-methylated CpG data around important transcription control elements. Transcriptional activity is controlled by the methylation status of these elements and accurately defining transcriptional control elements is therefore important.

Heatmaps generated by deepTools show the distribution of coverage in a 2 kb window around transcription start sites (TSS) (A), and a 1kb window around CTCF transcriptional repressor protein binding site (B), CpG islands (C) and H3K27me3 histone methylation site (D).

In all cases, EM-seq libraries have greater, and more uniform coverage than WGBS, and the enhanced coverage demonstrated using EM-seq results in fewer spurious methylation calls.

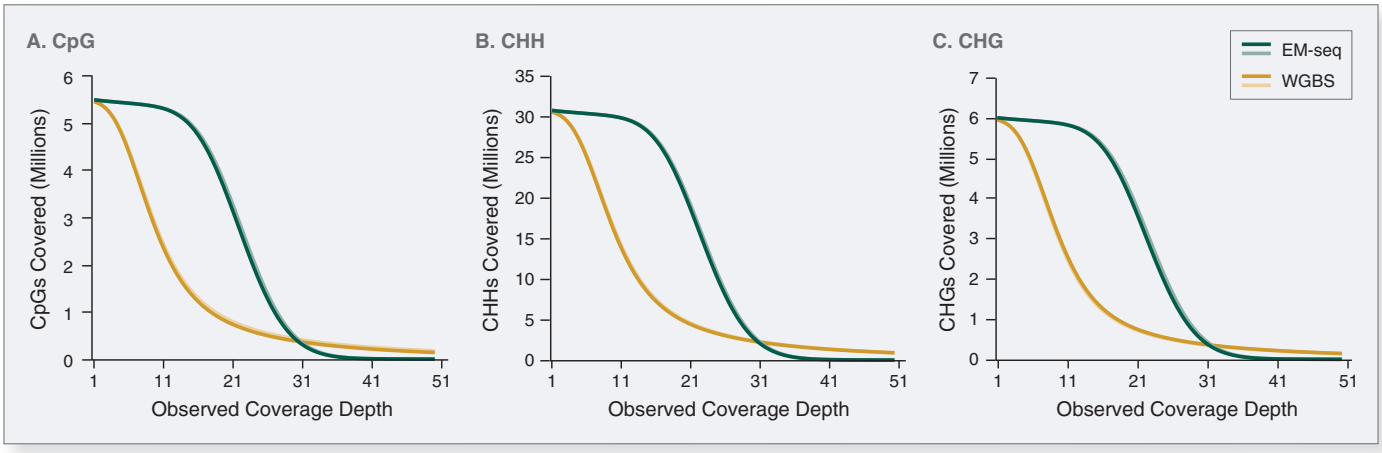
EM-seq is superior to WGBS for detection of CpG methylation around transcription start sites

Methylation of CpGs at each input was determined around the transcription start site (TSS). The 4 kb window around TSS was divided into 400 10 bp bins, and CpGs within these 10 bp bins with 8X or higher coverage were used to plot methylation. A: EM-seq has higher and more uniform coverage across TSSs . B: The average percentage methylation of 8X covered CpGs for EM-seq and WGBS libraries is shown. The EM-seq data is more representative of the expected methylation pattern across TSSs, with lowest levels at the TSS, and increasing methylation at the +/- 2 kb extremes.

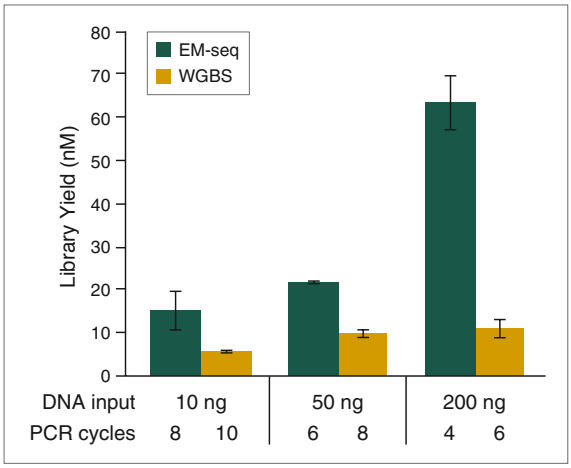


With *Arabidopsis thaliana*, EM-seq identifies more CpGs than WGBS, at lower sequencing coverage depth

50 ng *A. thaliana* genomic DNA was sheared to 300 bp using the Covaris S2 instrument and used as input into EM-seq and WGBS protocols. For WGBS, NEBNext Ultra II DNA was used for library construction, followed by the Zymo Research EZ DNA Methylation-Gold Kit for bisulfite conversion. Libraries were sequenced on an Illumina NextSeq 500 (2 x 75 bases). 125 million paired end reads for each library were aligned to TAIR10 using bwa-meth 0.2.2. CpG, CHH and CHG sites on both strands were counted independently. EM-seq identifies more CpGs, CHHs and CHGs, at higher coverage depth compared to WGBS, resulting in more usable information.

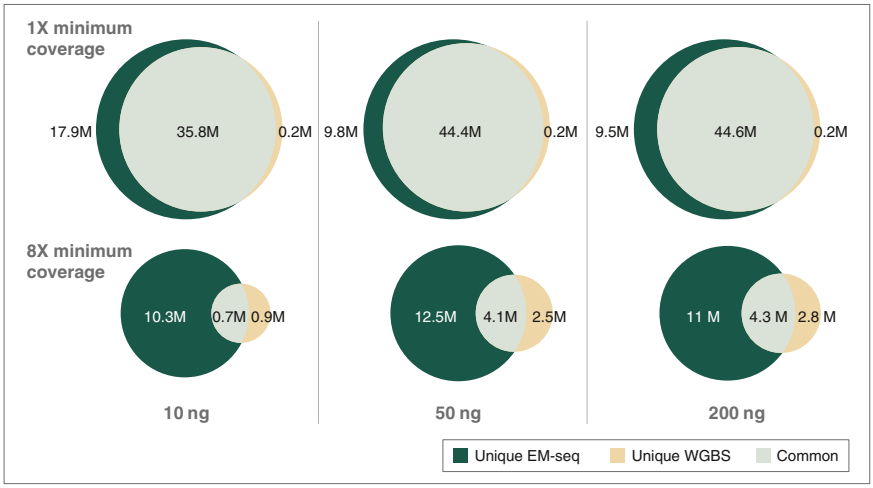


EM-seq produces higher yields than WGBS using fewer PCR cycles



10, 50 and 200 ng Human NA12878 genomic DNA was sheared to 300 bp using the Covaris® S2 instrument and used as input into EM-seq and WGBS protocols. For WGBS, NEBNext Ultra II DNA was used for library construction, followed by the Zymo Research EZ DNA Methylation-Gold™ Kit for bisulfite conversion. For all input amounts, EM-seq library yields were higher, and fewer PCR cycles were required, suggesting greater DNA loss in the WGBS protocol. Error bars indicate standard deviation.

EM-seq identifies more CpGs than WGBS, at lower sequencing coverage depth



Coverage of CpGs with EM-seq and WGBS libraries was analyzed using 324 million paired end reads. The number of unique and common CpGs identified by EM-seq and WGBS at 1x and 8x minimum coverage for each input amount are shown. EM-seq covers at least 20% more CpGs than WGBS at 1x minimum coverage threshold. The difference in CpG coverage increases to two-fold at 8x minimum coverage threshold.

Learn more and request a free sample at
www.NEBNext.com

ORDERING INFORMATION

PRODUCTS	NEB #	SIZE
NEBNext Enzymatic Methyl-seq Kit	E7120 S/L	24/96 rxns
NEBNext Enzymatic Methyl-seq Conversion Module	E7125S/L	24/96 rxns
NEBNext Q5U™ Master Mix	M0597S/L	24/96 rxns
NEBNext Multiplex Oligos for Enzymatic Methyl-seq (Unique Dual Index Primer Pairs)	E7140S/L	24/96 rxns

yNEB132 – 05/19

Your local NEB distributor:

BeNeLux:
BIOKÉ
Tel: (+31) 71 720 0220
(BE: 0800 - 71640)
Fax: (+31) 71 891 0019
info@bioke.com
www.bioke.com

BULGARIA
ELTA 90M Ltd.
Tel: (+359) 2 983 96 49
Fax: (+359) 2 9832211
office@elta90.eu
www.elta90.com

CROATIA:
DIAGNOSTICA SKALPELI D.O.O.
Free call: 800-315911
Tel: (+385) 1 377 8484
Fax: (+385) 1 377 8585
skalpeli@skalpeli.hr
www.skalpeli.hr

CZECH REPUBLIC:
BIOTECH A.S.
Tel: (Toll Free) 0800 124683
Fax: 02 72701742
info@ibiotech.cz
www.ibiotech.cz

DENMARK:
BioNordika Denmark A/S
DENMARK
Tel: (39) 56 20 00
Fax: (39) 56 19 42
info@bionordika.dk
www.bionordika.dk

FINLAND:
BioNordika Oy
Tel: (+358) 207 410 270
info@bionordika.fi
www.bionordika.fi

GREECE:
BIOLINE SCIENTIFIC
Tel: (0030) 210 522 6547
Fax: (0030) 210 524 4744
info@bioline.gr
www.bioline.gr

HUNGARY:
KVALITEX SCIENTIFIC
Phone: (1) 340 4700
Fax: (1) 339 8274
info@kvalitex.hu
www.kvalitex.hu

ITALY:
EUROCLONE S.P.A.
Free call: 800-315911
Tel: +39 02 38195.1
Fax: +39 02 3810145
info@euroclone.it
www.euroclonegroup.it

ISRAEL:
ORNAT LTD.
Tel: (+972) 8 947 7077
Fax: (+972) 8 936 3034
ornatbio@ornat.co.il
www.ornat.co.il

NORWAY:
BioNordika AS
Tel +47 23 03 58 00
Fax +47 23 03 58 01
info@bionordika.no
www.bionordika.no

POLAND:
Lab-JOT Ltd. Sp.z o.o. Sp.k.
Stacjonarny: +48 22 335 988 4
Komórkowy: +48 606 338 879
Fax: +48 22 335 981 9
biuro@labjot.com
www.labjot.com

PORTUGAL:
Werfen
Tel. +351 214 247 312
Fax +351 21 417 26 74
dac-pt@werfen.com
www.pt.werfen.com

ROMANIA:
ELTA 90 MEDICAL RESEARCH SRL
Tel: +40 21 2322694
Fax: +40 21 2322696
office@elta90mr.ro
www.elta90mr.ro

RUSSIA:
SkyGen LLC
Tel: +7 495 215 02 22
info@skygen.com
www.skygen.com

SERBIA:
ALFA GENETICS D.O.O.
Tel: 00381 11 3291219
Fax: 00381 11 2761281
office@alfagenetics.rs
www.alfagenetics.rs

SPAIN:
Werfen
Tel: +34 902 20 30 90
Fax: +34 902 22 33 66
customerservice-es@werfen.com
www.es.werfen.com

SLOVAK REPUBLIC
BioTech s.r.o.
Tel: +421 2 54 77 44 88
Fax: +421 2 32 20 10 48
info@ibiotech.sk
www.ibiotech.sk

SLOVENIA:
Mikro+Polo d.o.o.
Tel: +386 2 614 33 00
Fax: +386 2 614 33 20
info@mikro-polo.si
www.mikro-polo.com

SOUTH AFRICA:
INQABA BIOTEC
Tel: +27 123 43 58 29
Fax: +27 86 677 8409
info@inqababiotec.co.za
www.inqababiotec.co.za

SWEDEN, ESTONIA, LATVIA, LUTHUANIA:
BIONORDIKA SWEDEN AB
Tel: +46 8 30 60 10
Fax: +46 8 30 60 15
info@bionordika.se
www.bionordika.se

SWITZERLAND:
BIOCONCEPT
Tel: +41 (0)61 486 80 80
Fax: +41 (0)61 486 80 00
info@bioconcept.ch
www.bioconcept.ch

TURKEY:
EKA BIOLAB TEKNOLOJI
Tel. +90/312 248 13 24
Email: info@ekabiolab.com
www.ekabiolab.com

For a complete list of all the NEB distributors go to

www.neb.com

BIOKÉ
www.bioke.com