Research &
Pharma Solutions

Whole Genome Methylation Sequencing

Tech Note

Whole Genome Methylation Sequencing

An effective method to accurately analyze DNA methylation

DNA methylation is an important epigenetic mechanism that plays a key role in many different biological processes such as cellular development and differentiation, genomic imprinting, and regulation of tissue-specific gene expression. Genome-wide methylation analysis is usually performed using Whole Genome Bisulfite Sequencing (WGBS), a technique that is based on the chemical conversion of cytosines via bisulfite treatment followed by Next-Generation Sequencing (NGS). Although considered as the gold standard, WGBS has several drawbacks including DNA degradation and loss, which cause incomplete conversion and affect downstream analyses. EM-seq, a new enzyme-based conversion method claims to address these drawbacks and might represent a modern alternative for complete and efficient methylome analysis.

In this TechNote, we assess key quality parameters to evaluate the performance of the two NGS-based methylation assays WGBS and EM-seq using different sample types (genomic DNA, FFPE DNA, and cfDNA). Furthermore, we apply the EM-seq technique to 11 real-world samples (FFPE DNA) from glioblastoma patients and compare our results to the Illumina Infinium MethylationEPIC BeadChip array.

Methods

EM-seq (NEBNext® Enzymatic Methyl-seq, New England Biolabs) and WGBS (Accel-NGS® Methyl-Seq, Swift Biosciences) libraries were prepared using 100 ng of gDNA and 20 ng of FFPE DNA and cfDNA according to manufacturer's instructions. To assess repeatability of both assays with different sample types (gDNA, FFPE DNA, and cfDNA), all samples were analyzed in three replicates. For WGBS, cytosine conversion was performed prior to library preparation using EZ DNA Methylation Lightning™ Kit (Zymo Research). Treatment of DNA with bisulfite converts non-methylated cytosines into uracils (U) and subsequentially into thymines (T) during PCR, leaving methylated cytosines unaffected. For EM-seq, the enzymatical two-step conversion reaction takes place during library preparation. In the first reaction, TET2 oxidizes 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) to provide protection from deamination. In the second reaction, APOBEC is used to deaminate unmethylated cytosines into uracils. All libraries were paired-end sequenced on the Illumina NovaSeq 6000 instrument with a read length of 2 x 100 bp. Mapping and methylation calling was performed using the Illumina DRAGEN BIO-IT platform v 3.9.5. To exclude the effect of uneven sequencing depth between both assays, raw reads were downsampled to approximately 600 M reads (60 GB) for each sample. Further analyses were performed using R (Version 4.1.3). Top and bottom strand CpG dinucleotides were merged, yielding 28,306,459 detectable CpG sites. Correlations were plotted using methylKit package (Version 1.20.0).

Table 1: *Overview of assessed quality parameters. Values are means from gDNA, FFPE DNA and cfDNA triplicates.*

	gDNA		FFPE DNA		cfDNA	
	EM-seq	**WGBS**	**EM-seq**	**WGBS**	**EM-seq**	**WGBS**
Mapping efficiency (%)	83.9	80.6	75.6	72.9	77.7	79.9
Duplicates (%)	13.0	11.8	20.6	22.3	8.8	11.1
Mean insert size (bp)	339	180	202	144	158	155
Average CpG coverage	14.0	11.2	10.3	6.7	11.9	10.3

Results

Comparison of EM-seq and WGBS data

EM-seq and WGBS showed comparable duplication rates and mapping efficiencies across all sample types (table 1). However, EM-seq libraries were larger and showed slightly higher average CpG coverages than WGBS libraries. This indicates minimized DNA degradation during conversion and results in more usable reads and more accurate methylation analysis (table 1). Complete conversion is essential for accurate methylation analysis, as incomplete conversion results in overestimation of methylation. Thus, conversion efficiencies were examined for both assays by measuring methylation levels of two control DNAs (pUC19 and lambda), that were spiked to each sample. The pUC19 control DNA is fully methylated, which was correctly determined with both assays (figure 1A). The fully unmethylated lambda control DNA showed a methylation close to zero for EM-seq libraries and methylation values of approximately 2% for WGBS libraries.This indicates a slight methylation overestimation with the WGBS assay.

In the human genome, methylated cytosines predominantly occur in CpG dinucleotides. Methylation in non-CpG context is very uncommon and only occurs at low levels in specific human cell types (stem cells, neurons, glia cells) (Jang et al. 2017, Lister et al. 2013, Guo et al. 2014). We assessed global methylation levels across CpG and non-CpG (CHG and CHH) context for all sample types (figure 1B). Methylation values in CpG context were slightly higher for WGBS libraries compared to EM-seq libraries. In addition, and independent of the sample type, WGBS libraries also showed a constant methylation in CHG and CHH context of around 3%, indicating less accurate methylation detection. For EM-seq libraries generated from gDNA and cfDNA, methylation in non-CpG context was close to zero. Higher values were only observed for FFPE DNA (10% for CHG and 11% for CHH). However, this is not surprising, as the used FFPE DNA was derived from brain tumor tissue and lower levels of methylation in non-CpG context are expected in such material.

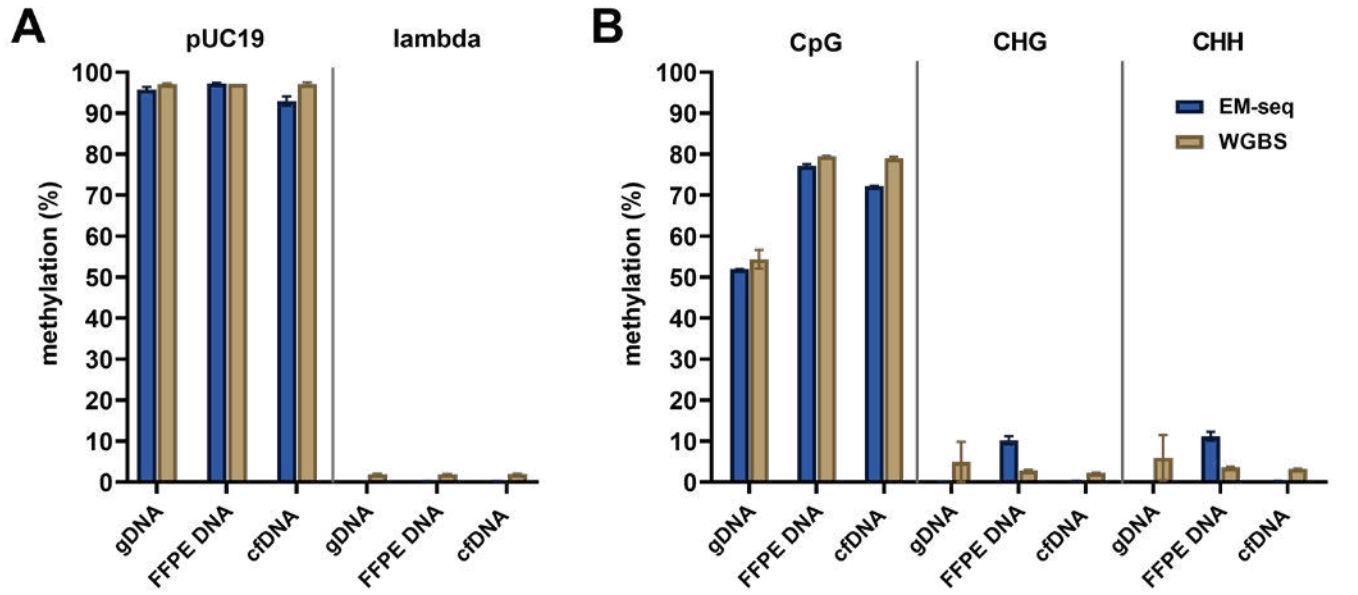

Figure 1: ***(A)*** *Conversion efficiency and* ***(B)*** *global methylation across CpG, CHG and CHH context determined for each sample type and both assays. Values are means from gDNA, FFPE DNA, and cfDNA triplicates.*

Table 2: *Percentage of detected CpG sites for each sample type using coverage thresholds 1x, 5x and 10x.*

	gDNA		FFPE DNA		cfDNA	
	EM-seq	WGBS	EM-seq	WGBS	EM-seq	WGBS
Coverage ≥1	97.5	96.9	97.3	94.4	96.6	96.0
Coverage ≥ 5	95.5	91.2	88.6	65.1	93.7	86.7
Coverage ≥ 10	82.7	63.0	53.1	23.7	67.9	55.3

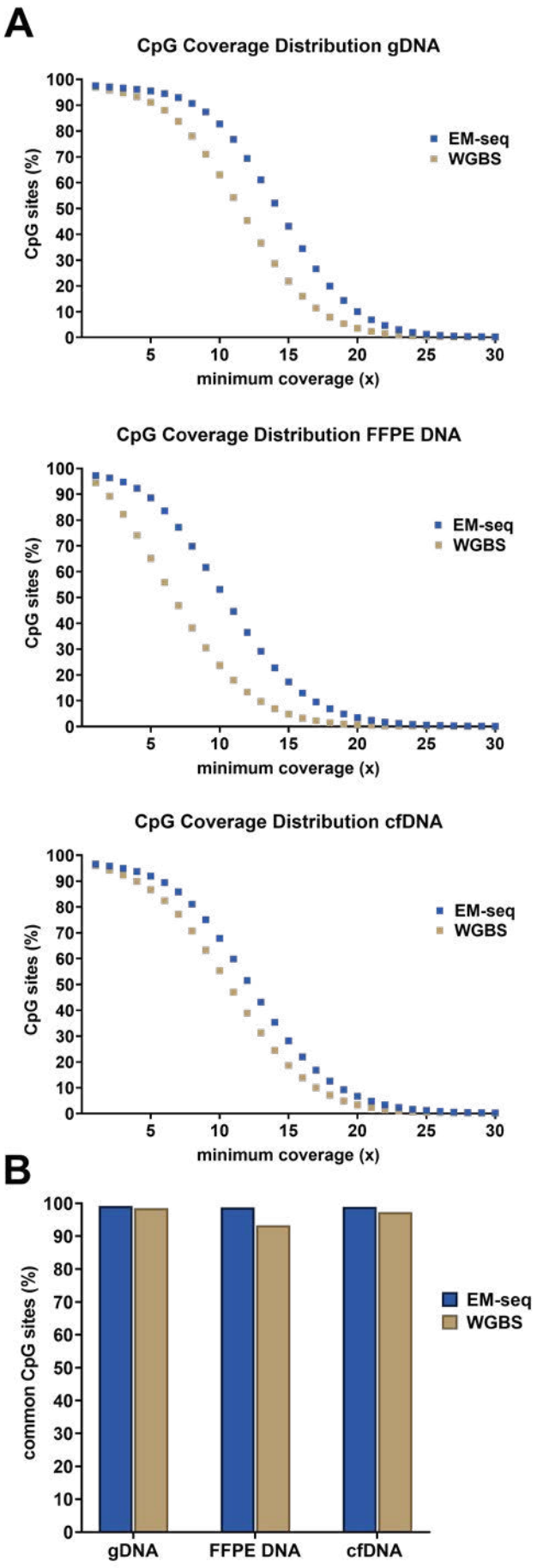

Figure 2: ***(A)*** *Percentage of covered CpG sites at increasing minimum coverage (1x-30x).* ***(B)*** *Percentage of commonly covered CpG sites between triplicates for each sample type.*

The human genome harbors approximately 28 million CpG sites. EM-seq covered 97% of these in all sample types and was closely followed by WGBS, which was able to cover approximately 96% of all CpG sites (table 2). When considering only CpG sites covered with at least 5x or 10x, we observed stronger differences between both kits, with EM-seq considerably outperforming WGBS (table 2). In addition, EM-seq covers more CpG sites at higher coverage than WGBS (figure 2A).

We also analyzed whether the detected CpG sites are commonly shared between replicates of a sample type. Our results revealed that approximately 99% of the CpG sites were commonly covered in all EM-seq replicates of a sample type at a coverage ≥ 1, indicating a higher repeatability in terms of CpG coverage for EM-seq compared to WGBS (figure 2B).

To further compare consistency of CpG methylation between replicates, Pearson correlations were computed using CpG sites that are covered with at least 10x in all replicates. We observed that methylation is more consistent between replicates from EM-seq than from WGBS libraries (figure 3).

For each sample type, we also compared the number of shared CpG sites between EM-seq and WGBS. We observed that both assays identified mostly the same CpG sites, resulting in very high numbers of commonly covered CpG sites (gDNA: 98.7%, FFPE DNA: 93.2%, cfDNA: 97.8%). CpG sites that were not shared between assays were covered more frequently with EM-seq (gDNA: 1.2%, FFPE DNA: 6.5%, cfDNA: 1.8%) than with WGBS (gDNA: 0.2%, FFPE DNA: 0.3%, cfDNA: 0.3%).

In addition, both assays showed highly concordant methylation values at individual CpG sites. Pearson correlations using all CpG sites with a mean coverage of at least 10x in EM-seq and WGBS revealed correlation coefficients above 0.91 for all sample types (gDNA: 0.98, FFPE DNA: 0.95, cfDNA: 0.92).

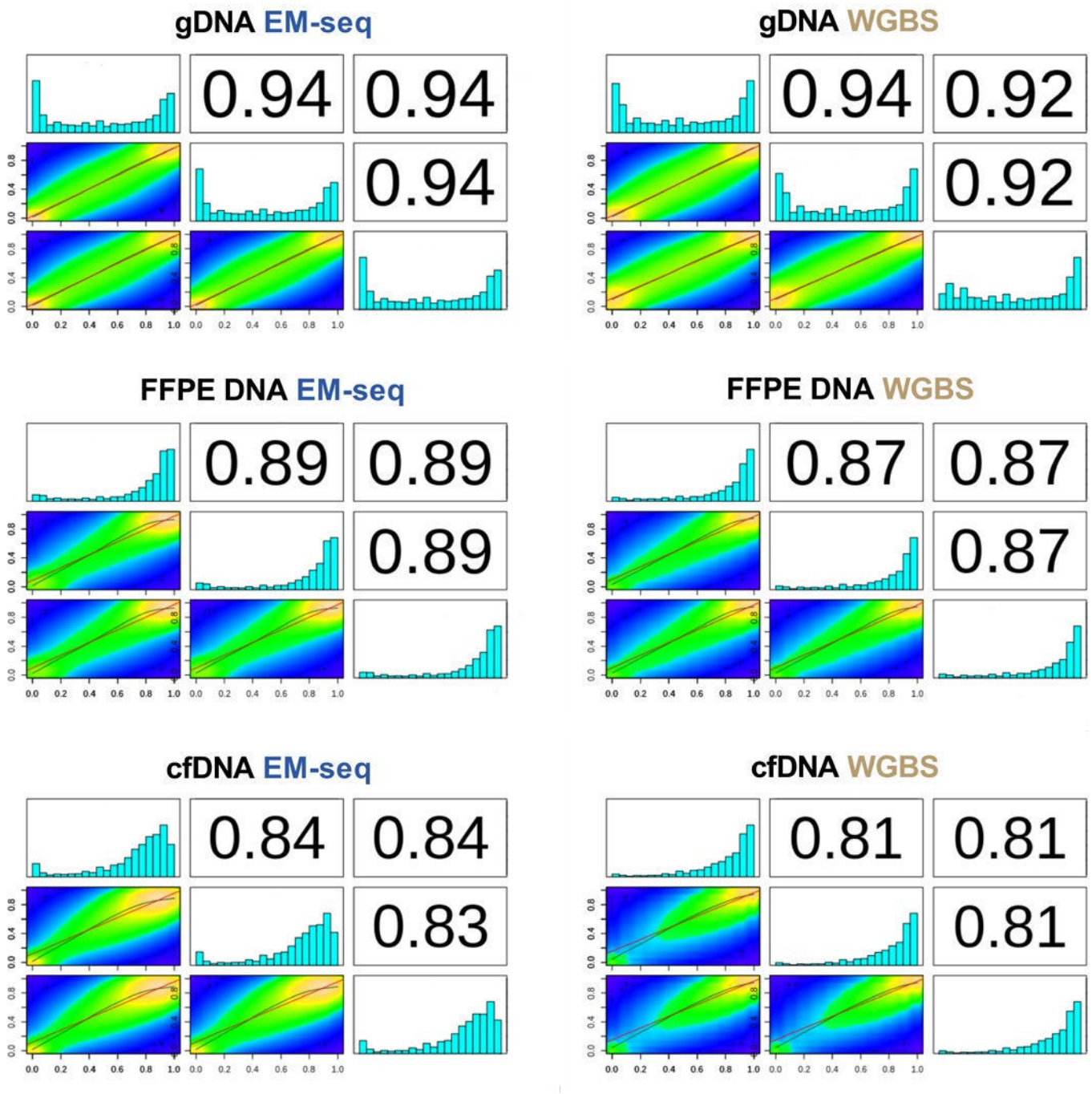

Figure 3: *Concordance in methylation between triplicates separated by sample type and assay (left EM-seq, right WGBS). Pearson correlations were computed for all CpG sites with a minimum coverage of 10. Values represent Pearson correlation coefficients (Pearson r).*

Comparison of EM-seq and array data

To compare the results of EM-seq with a reference method, we applied EM-Seq to 11 FFPE DNA samples from glioblastoma patients that have also been analyzed using Illumina's Infinium MethylationEPIC Beadchip array. For comparison, only CpG sites that are shared between both methods were considered (862,927 CpG sites). Both methods showed similar mean CpG methylation with high Pearson correlation coefficients and low differences in methylation at individual CpG sites (table 3, figure 4A).

In addition, EM-seq determined the same *MGMT* promoter methylation status as the array in 9 out of 11 patients (figure 4B). The two remaining patients (patient 1 and 10) were also not completely oppositely classified, instead one method detected an intermediate methylation and the respective other determined methylation values associated with a methylated or unmethylated status.

Table 3: *Pearson correlation coefficients (Pearson r) and mean CpG methylation for both methods in all patient samples.*

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10	Patient 11
Pearson correlation coefficient	0.94	0.94	0.94	0.95	0.94	0.94	0.93	0.95	0.95	0.92	0.89
Methylation EM-seq (%)	57.7	56.8	38.4	45.1	57.3	43.1	50.8	37.2	62.8	51.1	52.3
Methylation array (%)	57.9	57.1	40.9	47.5	57.8	44.6	52.7	40.1	61.3	51.6	51.9

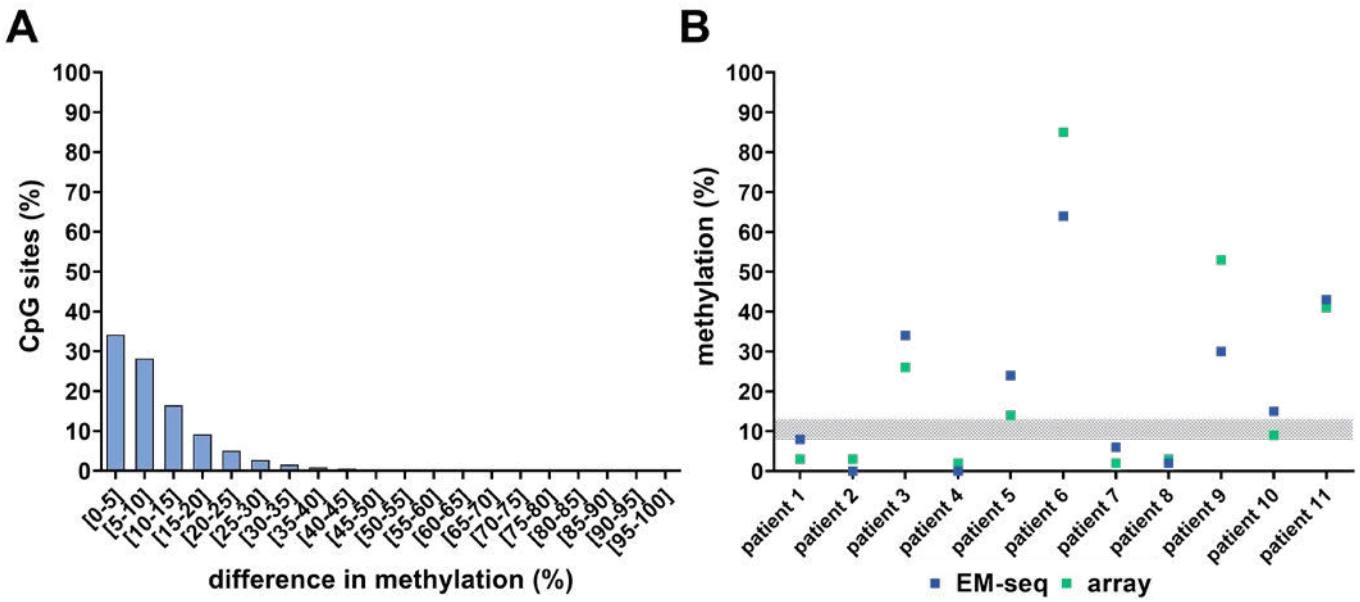

Figure 4: (A) *Differences in methylation between EM-seq and array. Methylation difference was calculated for each CpG site and subsequently assigned to its corresponding bin (x-axis).* ***(B)*** *Concordance of MGMT promoter methylation status classification between EM-seq and array. Classification was performed using the cutoffs described in Quillien et al. (2016), which is based on two CpG sites (cg12434587 and cg12981137), located in the MGMT promoter region (Bady et al., 2012). Methylation values < 8% were classified as "unmethylated", values between 8-13% as "intermediate", and values > 13% as "methylated". The grey shaded area represents the "intermediate area".*

Conclusion

Our comparison of two different NGS-based assays for genome-wide methylation analysis revealed that EM-seq yields considerably better results than WGBS for all tested sample types (gDNA, FFPE DNA, cfDNA). EM-seq produces larger and more intact libraries resulting in data with a highly uniform CpG coverage enabling accurate cytosine methylation analysis. In addition, EM-seq shows high repeatability in terms of coverage and methylation and its outcome is comparable to methylation arrays. Based on these results, EM-seq represents a very compelling alternative to WGBS, providing a particularly accurate analysis of genome-wide cytosine methylation at single-base resolution.

References

Bady P, Sciuscio D, Diserens AC, et al. MGMT methylation analysis of glioblastoma on the Infinium methylation BeadChip identifies two distinct CpG regions associated with gene silencing and outcome, yielding a prediction model for comparisons across datasets, tumor grades, and CIMP-status [published correction appears in Acta Neuropathol. 2013;126(1):159]. Acta Neuropathol. 2012;124(4):547-560. doi:10.1007/s00401-012-1016-2

Guo JU, Su Y, Shin JH, et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci. 2014;17(2):215-222. doi:10.1038/nn.3607

Jang HS, Shin WJ, Lee JE, Do JT. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. Genes (Basel). 2017;8(6):148. doi:10.3390/genes8060148

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, Ecker JR. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341(6146):1237905. Doi: 10.1126/science.1237905. PMID: 23828890; PMCID: PMC3785061.

Quillien V, Lavenu A, Ducray F, et al. Validation of the high-performance of pyrosequencing for clinical MGMT testing on a cohort of glioblastoma patients from a prospective dedicated multicentric trial. Oncotarget. 2016;7(38):61916-61929. doi:10.18632/oncotarget.11322

About Us

CeGaT GmbH is a global provider of genetic diagnostics and mutation-related disease analyses. The company combines its next-generation sequencing (NGS) process and analysis pipelines with its medical expertise – dedicated to identifying the genetic cause of disease and supporting patient management.

Genetic mutations can trigger a wide range of diseases, from epilepsy to Parkinson's. Through the use of NGS, it is possible to analyze all genes associated with a disease phenotype simultaneously – both fast and effectively. An interdisciplinary team of scientists and physicians evaluates the data and summarizes the findings in a comprehensive medical report. All services are performed in-house.

CeGaT, founded in 2009 and based in Tübingen, Germany, is accredited according to CAP, CLIA and DIN EN ISO/IEC 17025:2018.

CeGaT GmbH
Research & Pharma Solutions
Paul-Ehrlich-Str. 23
72076 Tübingen
Germany

Phone: +49 7071 56544-333
Fax: +49 7071 56544-56
Email: rps@cegat.com
Web: www.cegat.com

DAkkS Deutsche Akkreditierungsstelle D-PL-13206-01-00

Accredited by DAkkS according to DIN EN ISO/IEC 17025:2018

CAP ACCREDITED COLLEGE of AMERICAN PATHOLOGISTS

CLIA CERTIFIED ID: 99D2130225

2022/09