



NGS library preparation for balanced, comprehensive methylome coverage from low-input quantities

Next-generation sequencing (NGS) of bisulfite-converted DNA to detect methylation status with per-base resolution is currently restricted by the requirement of large amounts of input material. Here we show how the Accel-NGS[®] Methyl-Seq DNA library kit, with its unique chemistry, enables the construction of high-complexity libraries from 1 ng of input. With this kit, scientists will no longer be forced to sacrifice library quality when using low quantities of starting material.

Epigenetic research has been adversely affected by the amount of starting material required to perform whole-genome bisulfite sequencing (WGBS) of converted DNA samples. In addition to reducing the amount of required starting material by 100 times, Swift Biosciences has improved the overall genomic coverage across all inputs compared with two commercially available library preparation kits.

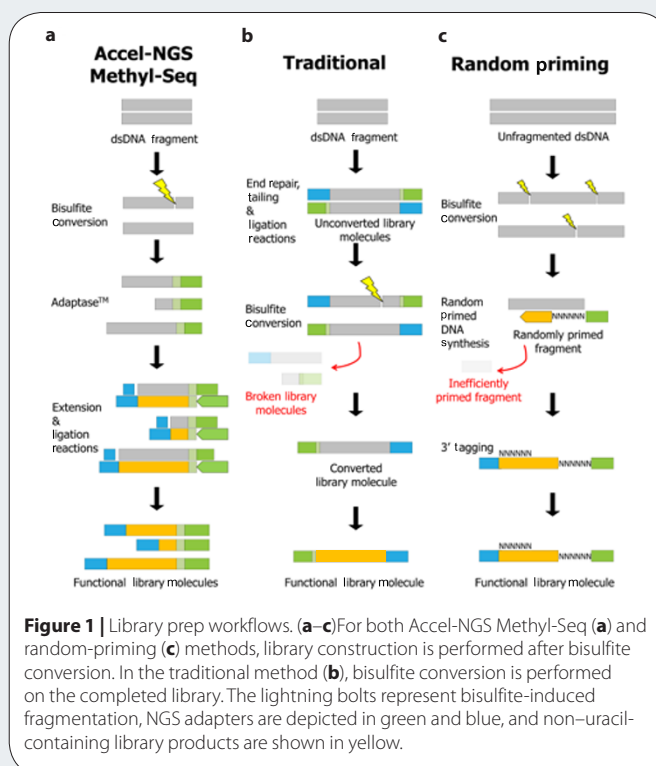
Accel-NGS Methyl-Seq is based on Swift Biosciences' Adaptase technology, which enables adapter attachment to single-stranded bisulfite-converted DNA (Fig. 1a). This preferable workflow eliminates the significant (up to 90%) library loss associated with workflows where the NGS library is constructed prior to bisulfite conversion (Fig. 1b). The template-independent adapter attachment of Accel-NGS Methyl-Seq also provides significant improvements in library quality compared with methods that incorporate adapters through random priming of single-stranded DNA (Fig. 1c).

Methods overview

The three library preparation methods mentioned above were evaluated using genomic DNA from *Arabidopsis thaliana*, a model organism for methylation analysis. Inputs of 100 ng, 10 ng or 1 ng *Arabidopsis* DNA were fragmented to an average of 350 bp by Covaris M220, if specified. Bisulfite conversion was performed using EZ DNA Methylation-Gold[™] according to the supplier's instructions. Duplicate libraries were constructed according to the suppliers' instructions and PCR amplified accordingly (four, seven and ten cycles; for traditional and random priming, the recommended cycling for 100 ng was used, with three additional cycles for each ten-times lower input). Libraries were sequenced on a HiSeq 2500 using V4 chemistry.

Cassie Schumacher, Laurie Kurihara & Kate Cunningham

Swift Biosciences, Inc., Ann Arbor, Michigan, USA. Correspondence should be addressed to C.S. (info@swiftbiosci.com).



All 18 samples were normalized to 30.2 million total reads and aligned to the *A. thaliana* (TAIR 10) reference genome using BSMAP for direct comparison of performance metrics.

An Accel-NGS Methyl-Seq library was also constructed using 10 ng of human HapMap DNA NA12878 from the Coriell Institute for Medical Research. This sample was processed in the same manner as the *Arabidopsis* samples, with the exception that 183.5 million reads were analyzed and aligned to the human (hg19) reference genome.

High-complexity library with Accel-NGS Methyl-Seq

Each method was tested at 100, 10 and 1 ng, all of which are within the specifications for the Swift Biosciences kit; the 10-ng and 1-ng

APPLICATION NOTES

Table 1 | Sequencing metrics

	Sample	Total reads	Reads aligned (%)	Genome coverage (x)	Duplicate reads (%)	Estimated library size	Relative library size
100 ng <i>Arabidopsis</i>	Accel-NGS Methyl-Seq	30.2 M	89.6	22	1.9	713,928,657	1
	Traditional	30.2 M	80.2	21	2.7	603,864,885	0.85
	Random primer	30.2 M	71.4	16	22.1	47,841,839	0.07
10 ng <i>Arabidopsis</i>	Accel-NGS Methyl-Seq	30.2 M	87.8	22	2.7	405,907,312	1
	Traditional	30.2 M	76.7	19	11.9	70,015,055	0.17
	Random primer	30.2 M	71.9	16	22.2	45,166,496	0.11
1 ng <i>Arabidopsis</i>	Accel-NGS Methyl-Seq	30.2 M	83.3	18	18.2	38,295,868	1
	Traditional	30.2 M	80.7	10	62.3	6,215,803	0.16
	Random primer	30.2 M	73.4	12	46.1	11,863,382	0.31
10 ng human	Accel-NGS Methyl-Seq	183.5 M	86.4	8.9	7.9	1,392,632,050	N/A

Each *Arabidopsis* file was normalized to 30.2 million reads, and data are reported as the average for duplicate bisulfite-converted samples. The relative library size is the ratio of the library size for alternative methods to that for Accel-NGS Methyl-Seq. For assessment of coverage for the human genome (bottom row), an Accel-NGS Methyl-Seq library was sequenced using HapMap NA12878 DNA.

values are below the specifications for the other methods. Accel-NGS Methyl-Seq yielded nearly 10% higher average alignment across the three inputs tested, with an average 87% alignment, compared with the traditional method's 79% and the random-priming method's 72% (Table 1). This gain in alignment data allows Accel-NGS Methyl-Seq to maximize the usable data generated from each sequencing run.

Duplicate reads can cripple methylation studies, as they must be removed before analysis can be performed. A proficient library preparation minimizes PCR duplicates by adapting input DNA molecules with high efficiency. With 10-times lower duplication than the random-priming method at inputs of 100 and 10 ng, and 3.4-times lower duplication than the traditional method at 1 ng, Accel-NGS Methyl-Seq powerfully demonstrated its efficiency in adapting bisulfite-converted fragments (Table 1). This gain in sequence data was also apparent in the library complexity as measured by the estimated library size (Table 1). Accel-NGS Methyl-Seq demonstrated the highest library complexity at 100 ng—more than ten times higher than that obtained with the random-priming method. Although all three methods had reduced estimated library sizes at 1 ng of input, the traditional method was the most sensitive, and its complexity fell six times lower than that of Accel-NGS Methyl-Seq. The gain in data output for Accel-NGS Methyl-Seq was also evident in the average genome coverage achieved (Table 1). It is important to note that these significant gains in data were achieved from the same number of total reads analyzed.

Accurate cytosine-methylation detection requires comprehensive CpX (CpG and CpH) coverage with balanced read depth across these genomic sequences; ideally, all unique CpX dinucleotides would be represented and covered at equal read depth. Accel-NGS Methyl-Seq also demonstrated the most balanced CpX coverage uniformity across all three input quantities, with a remarkable advantage at 1 ng of input (77% at ten-times coverage, compared with 17% and 31% for the traditional and random-priming methods, respectively) (Fig. 2a).

Human methylome coverage

Although *Arabidopsis* is a useful model organism, it lacks unique elements such as CpG islands that can be found in the human genome. Data generated from Accel-NGS Methyl-Seq with human DNA

demonstrated comprehensive coverage of unique CpG dinucleotide sequences, with only 2.2% missing with nine times the total sequencing coverage. Further, 96.4% of the methylome was covered at least five times (Fig. 2b). This lack of dropout and well-represented coverage from low-pass sequencing validates that Accel-NGS Methyl-Seq performs with great efficiency across different genomes, including those with a complex organization of base composition.

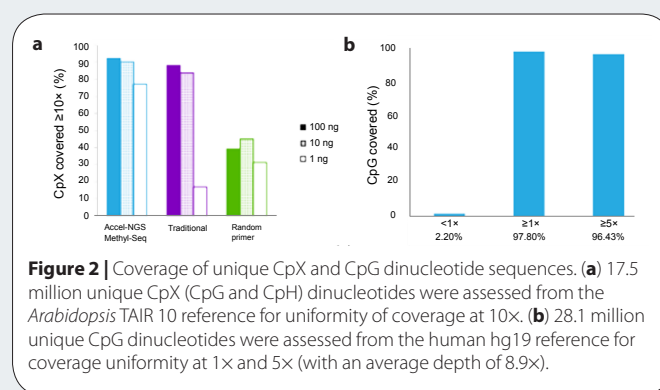


Figure 2 | Coverage of unique CpX and CpG dinucleotide sequences. (a) 17.5 million unique CpX (CpG and CpH) dinucleotides were assessed from the *Arabidopsis* TAIR 10 reference for uniformity of coverage at 10x. (b) 28.1 million unique CpG dinucleotides were assessed from the human hg19 reference for coverage uniformity at 1x and 5x (with an average depth of 8.9x).

In conclusion, Accel-NGS Methyl-Seq and the traditional method produced comparable data when the sample input was not limiting, but only Accel-NGS Methyl-Seq could preserve high library complexity at low sample inputs. Accel-NGS Methyl-Seq produced consistent data with minimal bias across all inputs, whereas the random-priming method delivered biased representation of the methylome at all sample inputs tested. The Accel-NGS Methyl-Seq kit from Swift Biosciences provided the most comprehensive, uniform methylome coverage from *Arabidopsis* and human samples, including those with low inputs. Further, the technology has been shown to maintain its performance with inputs from 100 pg to 100 ng when performing WGBS. This preservation of sample complexity empowers researchers to accurately characterize the methylome without large quantities of input material.

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.