APPLICATION NOTE

Accurate, Customizable, and Cost-effective Targeted Methylation Sequencing

ABSTRACT

Methylation sequencing is an important tool in understanding gene regulation during various processes, such as cell differentiation and disease progression, and is seeing increased use in diagnostic assays including oncology. Whole-genome methylation sequencing is ideal for target discovery with its unbiased and comprehensive single base resolution. Once specific genome regions of interest are identified, targeted methylation sequencing through hybridization capture prior to next generation sequencing (NGS) enables a more cost-effective methylation profiling solution that maximizes both the number of samples analyzed per experiment and the potential depth of sequencing for evaluating low frequency methylation signatures in complex samples (e.g., cell free DNA for liquid biopsy). In this application note, we demonstrate a highly accurate targeted methylation sequencing workflow using Swift Accel-NGS Methyl-Seq and Daicel Arbor Biosciences myBaits Custom Methyl-Seq systems. Through proprietary library preparation and hybridization chemistries, and a unique probe design algorithm, the system overcomes many inherent challenges to the efficient retrieval of specific bisulfite-converted DNA sequences. Using a target set comprising the promoter regions of over fifty cancer-related genes, we demonstrate that the workflow generates highly complex methylation sequencing libraries with low input quantity capabilities, and also accurately represents methylation profiles in a highly specific, unbiased fashion.

Key features include:

- Over 85% reads mapped to the genome following hybridization capture
- Over 80% of reads on-target, representing 8000 to 9000-fold enrichment
- Highly accurate genome-wide and site-specific methylation level measurement
- Compatible with as little as 1 ng starting DNA input

INTRODUCTION

DNA methylation occurs at cytosines within CpG and other dinucleotides and is a primary epigenetic mechanism controlling a variety of biological processes, especially transcriptional regulation of cell development and various diseases, including cancer. Whole genome bisulfite sequencing (WGBS) has become the gold standard in measuring methylation status genome-wide, as it enables methylation signatures to be discovered in a wide range of sample types [1]. Once regions of specific interest are identified and require characterization for large sample sets, WGBS becomes a relatively expensive option because of its comprehensive nature. Fortunately, hybridization capture, which can target thousands to millions of bases of specific genomic sequence in a single reaction, offers an extremely cost-effective solution that enables deeper sequencing and larger studies using a greater number of samples. Here we demonstrate a highly efficient, accurate, and reproducible targeted methylation sequencing workflow (Figure 1) that couples the Swift Accel-NGS Methyl-Seq library preparation with Daicel Arbor Biosciences myBaits Methyl-Seq hybridization system. We show how the system accurately reflects the region-wide as well as site-specific methylation levels obtained with standard whole-genome bisulfite sequencing, even in samples of mixed methylation states and low starting input.

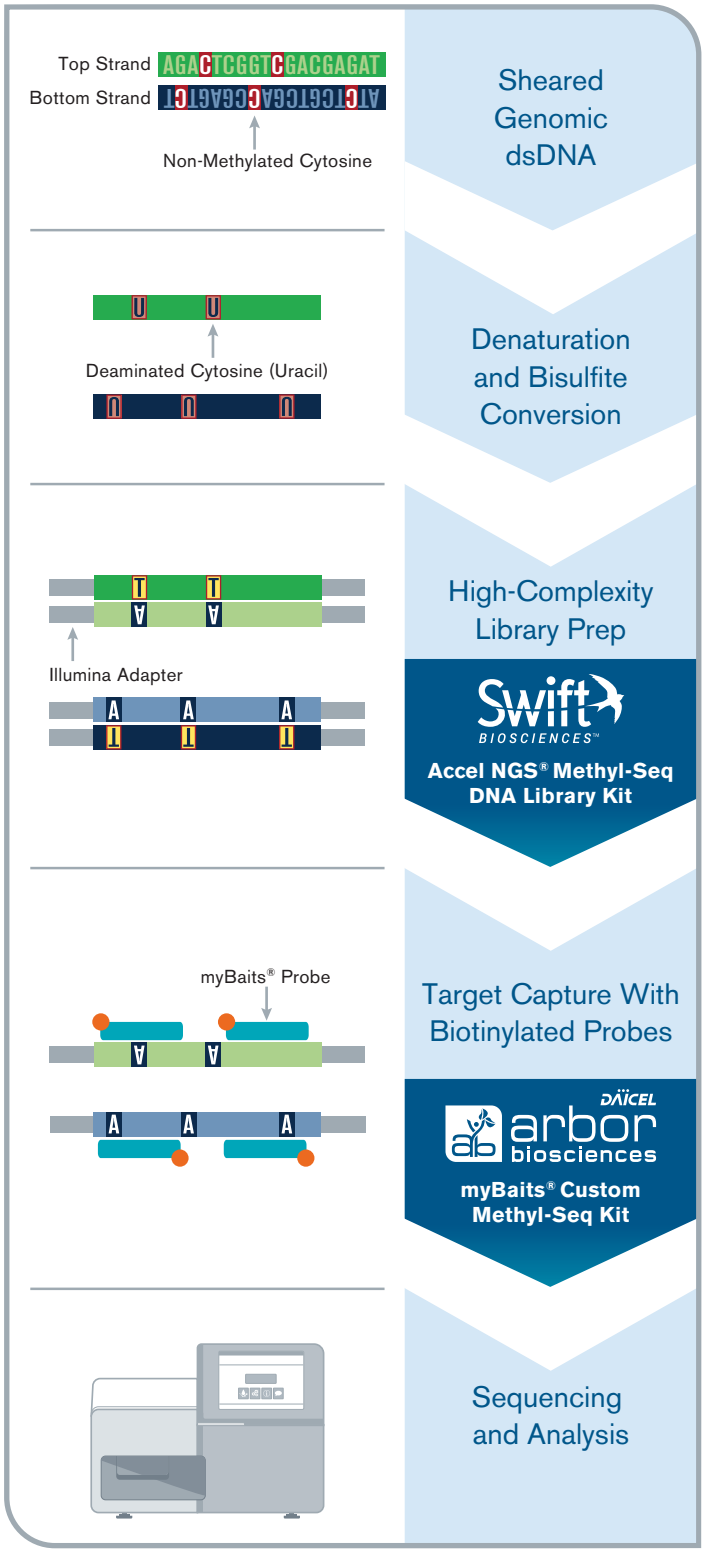

Figure 1. ***Simple and easy to implement.*** *Schematic illustration of the Swift Accel-NGS Methyl-Seq and Daicel Arbor Biosciences myBaits Custom Methyl-Seq target methylation sequencing workflow. Starting substrate is sheared genomic DNA with a specific set of methylation states (one of several potential states are depicted). After bisulfite conversion and library preparation, probes designed to accommodate a wide range of potential states are used in myBaits hybridization capture.*

SWIFT BIOSCIENCES

Accel-NGS Methyl-Seq Library Kit

The Swift Methyl-Seq kit is based on patented Adaptase® technology, which is a highly efficient and template-independent adapter attachment chemistry to single stranded DNA. This enables the construction of high complexity libraries from single stranded bisulfite converted DNA to provide comprehensive genome coverage, and can support a range of samples including liquid biopsy/cfDNA often limited to low inputs (<10 ng). Because adapter titration is not required, high ligation efficiency is maintained at low input quantities for the Swift kit. Methods that ligate methylated adapters followed by bisulfite conversion lose library complexity due to bisulfite induced fragmentation of the completed libraries that limits support of low input applications, and methods utilizing enzymatic conversion also cannot support inputs <10 ng.

DAICEL ARBOR BIOSCIENCES

myBaits Custom Methyl-Seq System

With myBaits Custom Methyl-Seq, Daicel Arbor Biosciences has adapted its foundational myBaits technology to meet the challenges of capturing cytosine-deaminated templates using pools of biotinylated RNA probes to hybridize genomic regions of interest. Traditional hybridization capture can tolerate several probe-target mismatches and still efficiently enrich target molecules. However, hybrid capture of cytosine-deaminated templates of unknown methylation state is faced with several unique challenges to the design and execution of the experiment:

1) for a given target sequence, more probe sequences are required to accommodate a wide range of possible methylation states among the cytosines,

2) GC content and thus probe-target hybridization strength is significantly lower after conversion of many cytosines, and

3) genomic sequence complexity is also significantly lowered following cytosine deamination, reducing the mean specificity of probes to their intended targets.

The points above magnify the costs of both probe manufacturing and sequencing, due in large part to overall lower potential specificity of a given probe design. The myBaits Custom Methyl-Seq system solves these problems in two ways. First, it uses a methylation-specific probe design algorithm that simulates some but not all potential cytosine methylation configurations on both strands from the converted genomic templates; probes are also aggressively filtered for specificity to both strands of a fully CpG-methylated genome sequence simulation. Additionally, the myBaits Custom Methyl-Seq system employs optimized target enrichment parameters to achieve high reads on-target rate with minimal locus dropout.

MATERIAL AND METHODS

To demonstrate how the Swift-Daicel Arbor targeted methylation sequencing system successfully reproduces methylation states at multiple genomic loci, we selected the putative promoter regions of 50 cancer-associated genes (total target size ~100 kb) as initial targets. A total of 20,003 probes were designed and synthesized for these regions using the Daicel Arbor Biosciences myBaits Custom Methyl-Seq probe design and filtration pipeline. Hereafter we focus our analysis on the regions probed by the design, which covers more than 92% of the initial intended promoter region space.

To address whether the system is compatible with a range of DNA inputs and methylation states, we coupled the myBaits kit with a number of sequencing libraries prepared with the Swift Methyl-Seq Library Preparation Kit. The first series of libraries used a range of bisulfite converted input DNA amounts (1 ng, 5 ng, and 100 ng) isolated from the NA12878 cell line (Coriell Institute). Extensive WGBS data is readily available for this cell line using the Swift kit, allowing us to determine the accuracy of our site-specific methylation calls following target enrichment. The second set of libraries comprised 50 ng total input from different combinations of bisulfite-converted fully-methylated and fully-nonmethylated DNA (Zymo Cat. Nos. D5014-2 and D5014-1), simulating approximately 10%, 30%, 50%, 70%, and 90% genome-wide methylation levels (referred to simulated libraries hereinafter). For library preparation, samples were sheared to 350 bp by Covaris M220, followed by bisulfite conversion using Zymo EZ DNA Methylation-Gold (Cat. No. D5006). This kit converts unmethylated cytosines to uracils and results in single-stranded DNA further fragmented to 170-200 bp. Accel-NGS Methyl Seq libraries were prepared according to manufacturer's recommendations, and for compatibility with hybridization capture, amplified using KAPA HiFi HotStart ReadyMix (Cat. No. KK2602) for both pre- and post- hybridization capture library amplification steps. A representative library from each simulated level was sequenced with PE150 sequencing on a partial NovaSeq S4 flowcell to directly measure the methylation level and evaluate reproducibility.

Target enrichment reactions used the myBaits version 5.0 High Sensitivity Protocol using two rounds of 63°C hybridization and wash temperatures. Libraries were captured in singleplex using 400 ng input per enrichment reaction, and in multiplex using 200ng input per sample and 8 samples per reaction. Following capture, libraries were pooled in equimolar ratio and sequenced on one Illumina MiSeq® lane using a 2 × 76 bp protocol with dual 8 nt indexing reads.

Raw reads were trimmed of adapter sequence and low-quality bases using *Trimmomatic* [2], and 15 nt from the 5' end of Read 2 were trimmed with *cutadapt* [3] as recommended to eliminate the majority of Adaptase tail sequence. Processed reads were aligned to the human genome (hg19) using *bismark* [4] with default parameters. PCR duplicates were removed using *deduplicate_bismark*, and methylation call and CpG coverage were calculated with *bismark_methylation_extractor*. We imported the methylation call files into R using *methylKit* [5] and used coverage of 5 as a filtering parameter. The Pearson correlation coefficient was then calculated among replicates and between WGBS and enriched libraries in *methylKit*.

RESULTS

High Coverage with Minimal Sequencing

To evaluate the hybridization capture performance, we assessed key metrics including mapping rate, percentage of reads on-target (specificity), and per-site depth of coverage in a random subsample of 950K raw sequencing reads for all samples. For NA12878, we obtained an average ~85% unique mapping rate to the genome, and an average of 84.6%, 84.0% and 76.7% raw reads on-target following enrichment of libraries with 100 ng, 5 ng and 1 ng DNA input respectively (Figure 2A). Compared to the 0.009% on-target rate measured in the WGBS data for this sample, this translates to roughly 8400-fold to 9300-fold enrichment. The mapping rate and the reads on-target rate did not change significantly with reduced DNA input, indicating input-independent high specificity of the system.

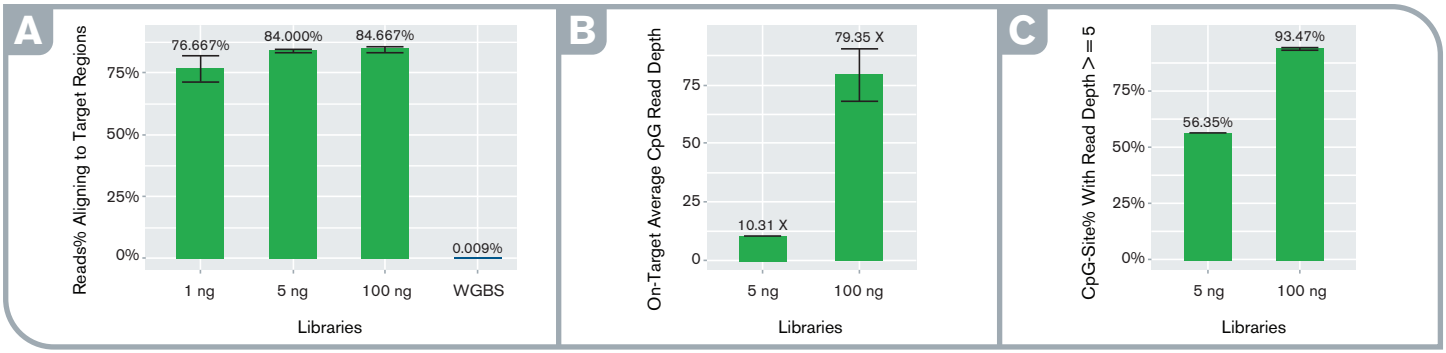

Figure 2. ***High specificity and orders of magnitude target enrichment.*** *Performance metrics: (A) percentage of on-target reads in enriched and WGBS libraries, (B) average per-site depth of coverage at on-target CpG sites at an even subsample of 950K read-pairs in the enriched libraries, (C) percentage of on-target CpGs with read depth >= 5 at the even subsample in the enriched libraries. Libraries were prepared with the Swift Accel-NGS Methyl-Seq Library Kit with varying input amounts and captured with the Daicel Arbor Biosciences myBaits Custom Methyl-Seq system.*

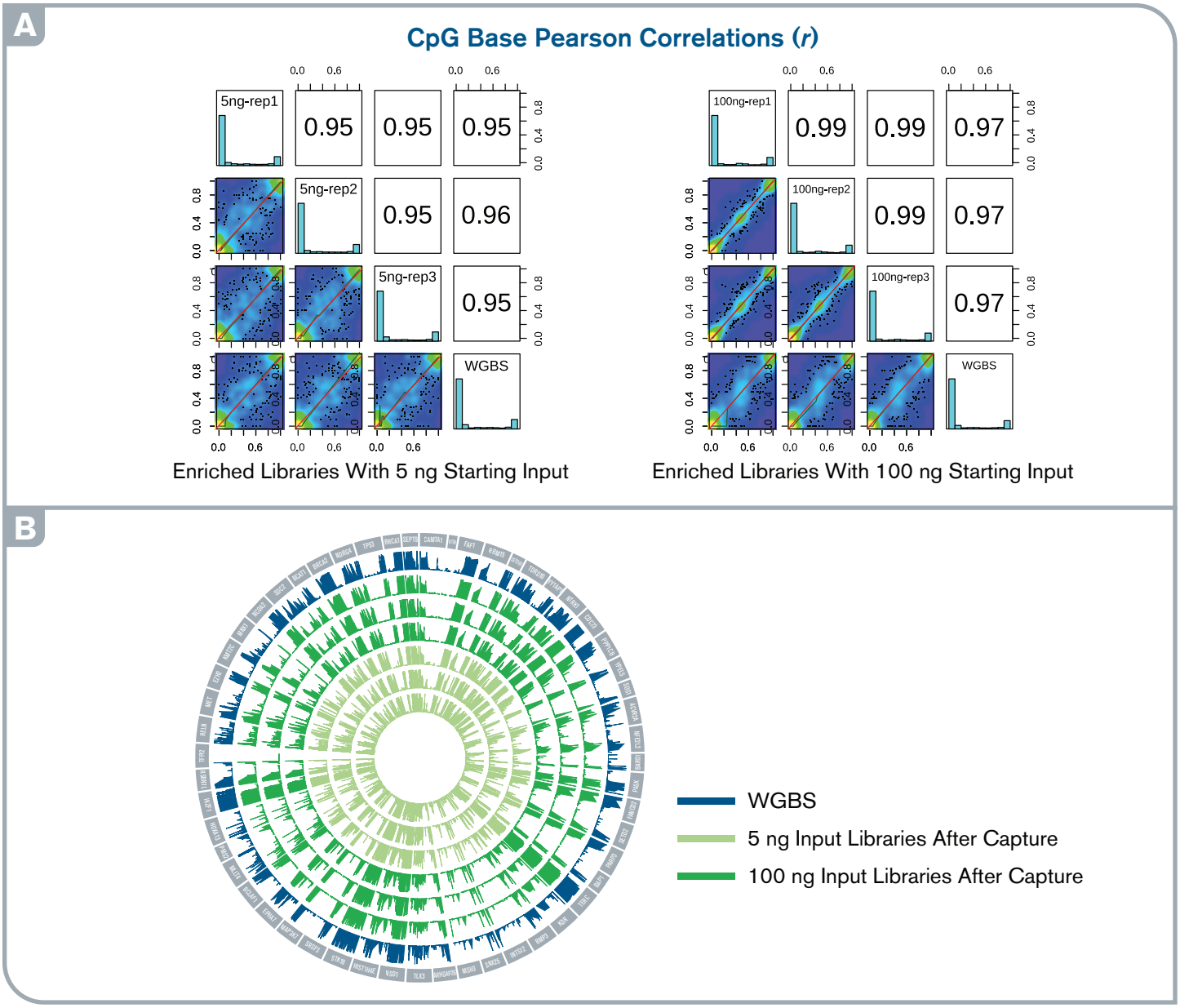

Figure 3. ***High accuracy and reproducibility of targeted methylation capture.*** *(A) Histograms and scatter plots show the comparison of CpG methylation levels across the target regions between WGBS and capture libraries and among replicates. (B) Circos plot illustrating methylation patterns for the entire target regions in capture and WGBS libraries.*

We achieved an average of 201× and 18× unique read coverage across all targeted sites, and an average of 79× and 10× read coverage on targeted CpG sites in the capture libraries from 100 ng and 5 ng DNA input respectively (Figure 2B). This enabled methylation states to be called at over 57% and 90% of the target region CpG sites with read depth >= 5 for libraries from 5 ng and 100 ng input, respectively (Figure 2C). This indicates that on a single MiSeq run, a target of this size can be resolved in at least 24 samples of comparable starting quality.

Highly Accurate Site-Specific Methylation Quantification

To assess the accuracy of our target methylation sequencing, we compared the distribution of methylation levels across the target regions between the enriched NA12878 libraries and WGBS data for this same cell line, both prepared with the Swift Methyl-Seq kit. This revealed high correlations of site-specific methylation quantification between WGBS and capture libraries regardless of input amount (Pearson correlation coefficient $r >= 0.95$ for 5 ng, and $r >= 0.97$ for 100 ng, Figure 3A). Remarkable correlations ($r >= 0.95$ for 5 ng, and $r >= 0.99$ for 100 ng, Figure 3A) were also observed among technical replicates. These strong correlations manifest in high correspondence in regional methylation calls, as illustrated in Figure 3B. Taken together, this indicates that the Swift Accel-NGS Methyl-Seq and Daicel Arbor Biosciences myBaits Custom Methyl-Seq system accurately reproduces the methylation measurements of WGBS for a fraction of the required sequencing depth.

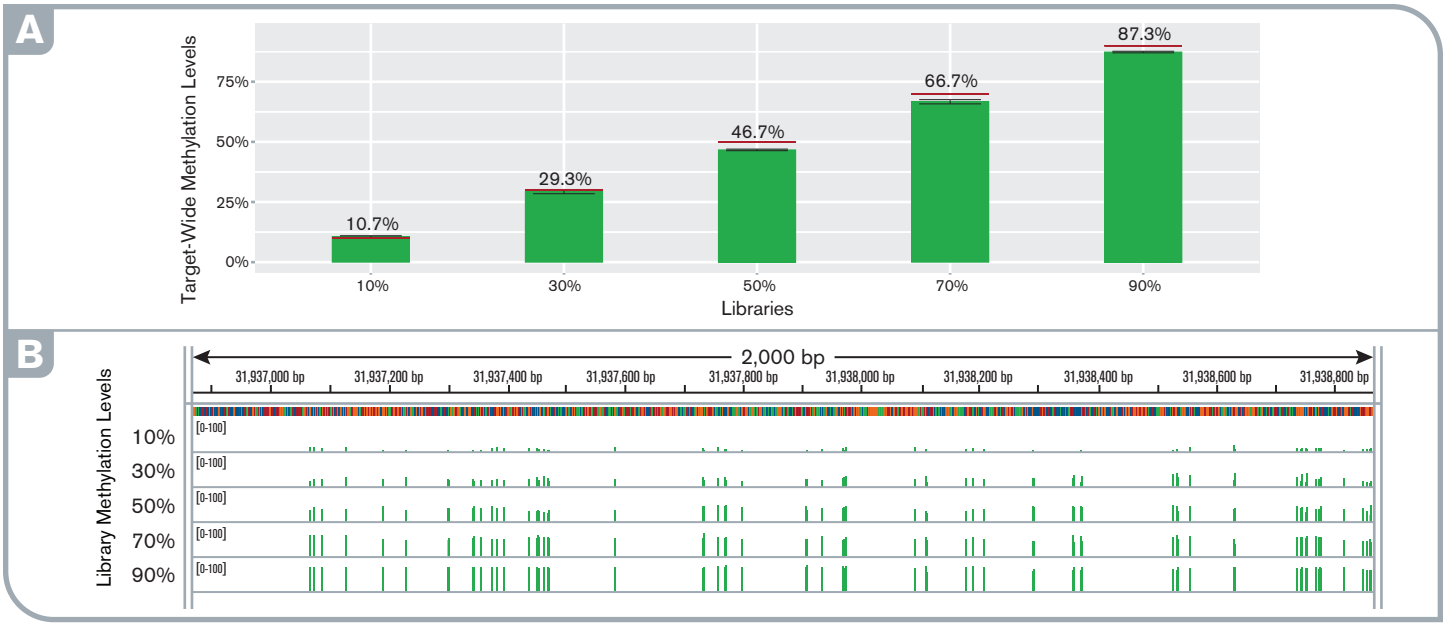

Figure 4. ***High sensitivity and fidelity of targeted methylation detection at any methylation levels.*** *(A) Target-wide methylation levels in capture libraries. Red lines indicate the simulated level. (B) IGV screenshot showing the uniform distribution of methylation levels along the gene STK12 promoter locus observed with the capture libraries at various methylation levels.*

High Sensitivity and Fidelity of Methylation Detection at All Levels

Many biological samples contain mixtures of cells or cell free DNA with different methylation states at various sites in their genomes. Therefore, it is critical that a hybridization capture system for methylation sequencing can reproduce these ratios accurately. To evaluate this with the Swift-Daicel Arbor system following enrichment, we measured target-wide methylation levels for the simulated libraries and found that these closely reflect the corresponding intended simulated levels (i.e., 10 to 90% methylated) (Figure 4A). Methylation levels at sites along the entire target region for all enriched libraries were also extremely consistent (Figure 4B), indicating little bias towards methylated or non-methylated states in the mixed samples on per-site basis.

Efficient Multiplexing Capabilities

Hybridization capture is performed on indexed sequencing libraries, and multiple libraries can be pooled into single enrichment reactions to enable even more cost-savings and increased throughput. To demonstrate that pooling libraries prior to enrichment does not compromise assay performance, we compared results obtained for libraries when pooled together prior to enrichment ("multiplex" 8 libraries per reaction) to the same results when enriched individually ("singleplex"). We observed no significant differences in sensitivity, efficiency, or accuracy between the multiplex and singleplex experiments (Figure 5), thereby demonstrating that the Swift-Daicel Arbor system has tremendous scalability potential.

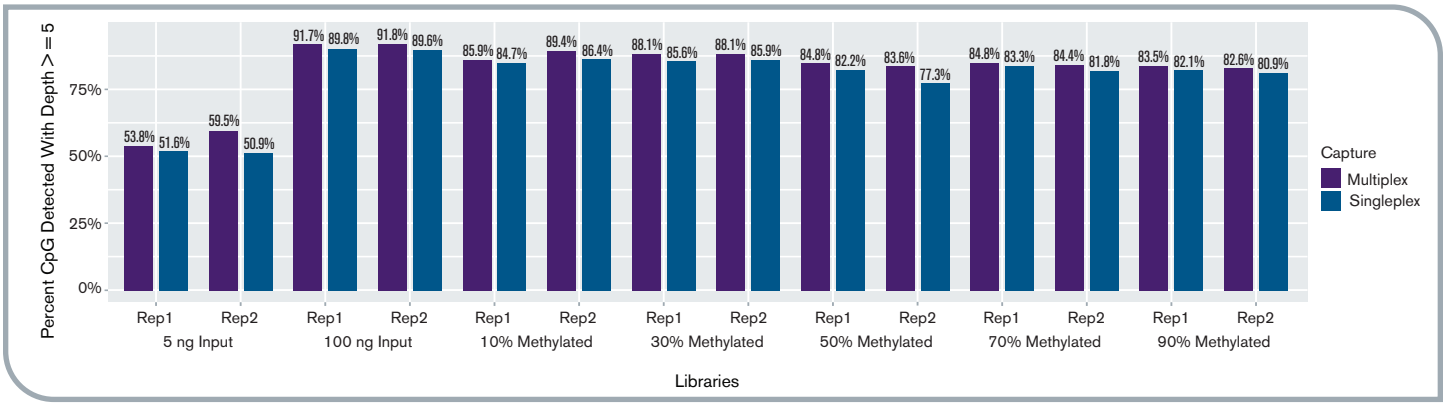

Figure 5. ***Comparable performance when enriching pools of libraries.*** *Percentage of target CpGs at 5× or higher read depth in a 460K raw read-pair subsample per library when captured in pools of 8 (Multiplex) vs. individually (Singleplex).*

PRODUCT TABLE

Cat. No.	Description	Reactions
30XX16M.v5	Daicel Arbor Biosciences myBaits® Custom Methyl-Seq Kit	16
30XX48M.v5	Daicel Arbor Biosciences myBaits® Custom Methyl-Seq Kit	48
30XX96M.v5	Daicel Arbor Biosciences myBaits® Custom Methyl-Seq Kit	96
30024	Swift Biosciences Accel-NGS® Methyl-Seq Library Prep Kit*	24
30096	Swift Biosciences Accel-NGS® Methyl-Seq Library Prep Kit*	96

** Swift Methyl-Seq Indexing Kits complete the library preparation workflow.*

CONCLUSION

The unique coupling of the highly robust Swift Accel-NGS Methyl-Seq and optimized design features of the Daicel Arbor Biosciences myBaits Custom Methyl-Seq systems provides an accurate and cost-effective targeted methylation sequencing solution. For a panel of 50 cancer-associated gene promoters, we show that the system allows the researcher to measure target DNA methylation levels even for low input DNA samples with a mixture of methylation states, and achieve remarkable specificity and sensitivity. This makes the Swift Accel-NGS Methyl-Seq and Daicel Arbor Biosciences myBaits Custom Methyl-Seq system a key tool for epigenetic research of any application, with particularly excellent potential in cancer-associated prognostic and biomarker discovery no matter the quality or quantity of starting nucleic acid material.

REFERENCES

1. Tost, J. ed. (2018) **DNA Methylation Protocols.** *Humana Press.*
2. Bolger, A.M., Lohse, M. and Usadel, B. (2014) **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics*, 30(15), pp. 2114-2120.
3. Martin, M. (2011) **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet. Journal*, 17(1), pp.10-12.
4. Krueger, F. and Andrews, S.R. (2011) **Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications.** *Bioinformatics*, 27(11), pp.1571-1572.
5. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A. and Mason, C.E. (2012) **methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles.** *Genome Biology,* 13(10), pp.1-9.

DAICEL arbor biosciences

web: www.arborbiosci.com
email: info@arbor.daicel.com
phone: 1-734-998-0751
twitter: @ArborBio