Cost-effective targeted methyl-seq using an xGen™ Custom Hyb Capture Panel and the xGen Methyl-Seq DNA Library Prep Kit identifies DNA methylation in low input samples consistent with WGBS

Performing comprehensive methylation profiling, such as whole genome bisulfite sequencing (WGBS), is a gold standard for target discovery. Once target methylation signatures are known, a more cost effective approach is desirable to increase the number of samples analyzed. This can be achieved with targeted methylation sequencing that is limited to methylation signatures of epigenetic interest. Using the xGen Methyl-Seq DNA Library Prep Kit paired with an xGen Custom Hyb Capture Panel to create target-enriched NGS libraries has been shown to improve coverage for desired regions at a significantly lower cost than WGBS without impacting data quality. The complete workflow supports methylation detection, high-performance target enrichment, and enables methylation detection in low input samples that is equivalent to data generated by WGBS.

Introduction

Changes in DNA methylation result in epigenetic modifications which affect gene regulation in biological processes, mammalian development, cell imprinting, and diseases. A common form of DNA methylation, 5' methylcytosine (5mC), occurs on CpG dinucleotides, which are in promoter regions of key genes involved in development and tissue specific gene regulation. Since methylation status plays a role in regulating gene expression and stabilizing gene silencing, it is a critical cancer biomarker in oncology research. Alterations in methylation patterns appear in early stages of cancer [1] and differences in CpG methylation status have been shown to distinguish normal tissue from cancer tissue [2]. Although using NGS in epigenetics research identifies changes in methylation patterns in samples such as formalin-fixed paraffin-embedded (FFPE), circulating tumor DNA (ctDNA) has also become a promising tool for minimally invasive longitudinal oncology research studies. Unfortunately, ctDNA represents only a fraction of the plasma-derived circulating cell-free DNA (cfDNA), therefore, strategies that are sensitive enough to identify rare transcripts in a heterogenous population of nucleic acid fragments from different cell types is required [3].

Whole genome bisulfite sequencing (WGBS) enables global analysis of methylation status and is the gold standard for biomarker discovery [4], however, the amount of sequencing required to assess the methylation status of all CpGs in a genome-wide analysis can incur higher costs. A targeted methylation sequencing approach can be used to study biologically relevant regions of interest at a higher coverage depth while reducing sequencing costs and increasing throughput capabilities [5]. The xGen Methyl-Seq DNA Library Prep Kit paired with an xGen Custom Hyb Capture Panel provides a cost-effective approach to generate high-quality enriched methyl-seq libraries for targeting specific loci of interest for methylation signatures.

> SEE WHAT MORE WE CAN DO FOR **YOU** AT **WWW.IDTDNA.COM**.

IDT INTEGRATED DNA TECHNOLOGIES

Workflow of the xGen Methyl-Seq Library Prep Kit followed by target capture with an xGen Custom Hyb Panel

Bisulfite conversion is a chemical process that involves treating denatured DNA with bisulfite [6]. The xGen Methyl-Seq DNA Library Prep Kit has a post-bisulfite library preparation workflow to maximize library complexity by converting bisulfite-converted single-stranded fragments directly into library molecules (Figure 1A). Traditional methods construct libraries from dsDNA using methylated adapters followed by bisulfite conversion, which leads to significant library loss due to bisulfite-induced DNA fragmentation (Figure 1B). The random priming methods implement bisulfite treatment first followed by random priming on the DNA fragment, which provides a reduced library complexity due to a lower efficiency biased workflow (Figure 1C). The xGen Methyl-Seq Library Prep Kit utilizes proprietary Adaptase™ technology that enables library preparation from bisulfite-converted single-stranded DNA (ssDNA) molecules in a highly efficient and template-independent manner. Using bisulfite-converted ssDNA molecules as library input provides an efficient workflow using low DNA input quantities while maintaining high library complexity and enables the preparation of libraries from a broader range of sample types and input amounts.

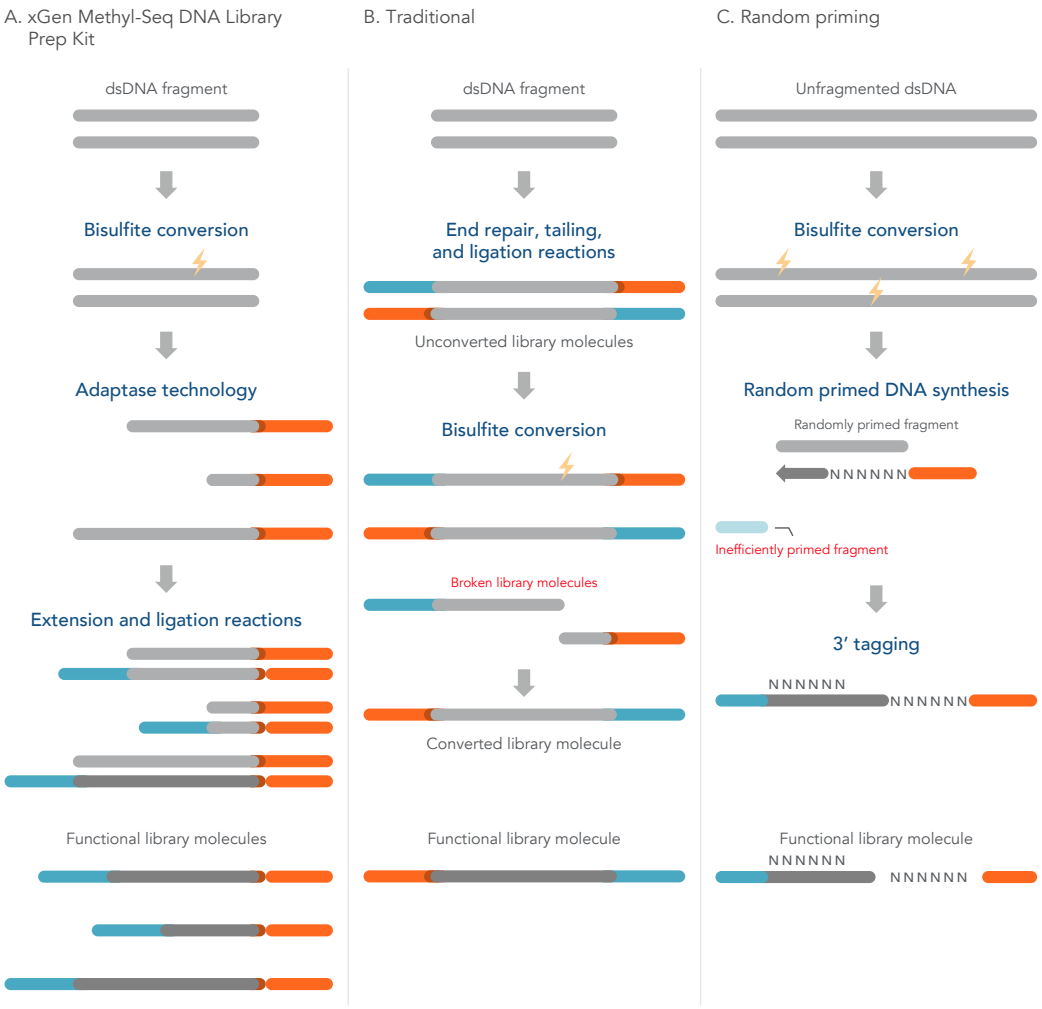

Figure 1. Comparison of the xGen Methyl-Seq DNA Library Prep Kit workflow to the other methods. (**A**) Before using the xGen Methyl-Seq DNA Library Prep Kit, double-stranded DNA (dsDNA) fragments undergo bisulfite conversion. The kit then directly converts these single-stranded fragments into library molecules with the proprietary Adaptase technology. (**B**) The traditional method of methylation sequencing builds the library before bisulfite conversion, which can lead to broken and lost library molecules. (**C**) The random priming method converts unfragmented dsDNA with bisulfite treatment first, but due to inefficient random priming, many library molecules are lost and are not represented in the final library.

The xGen Hybridization Capture workflow is a target enrichment solution for increasing the sequencing coverage in regions of interest for NGS libraries (Figure 2). The workflow provides researchers the flexibility to perform either a 4- or 16-hour hybridization capture and uses xGen Universal Blockers to significantly reduce off-target sequencing reads and enable the enrichment of multiple methyl-seq libraries in a single capture, which can reduce reagent costs and minimize hands-on-time—without sacrificing results.

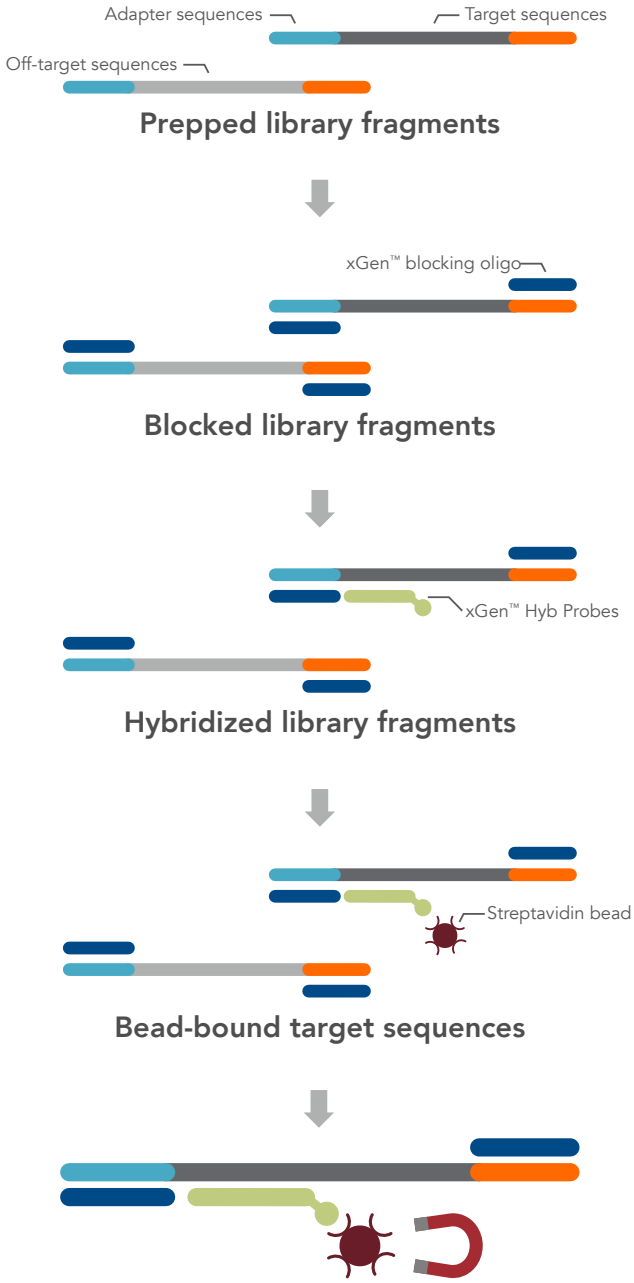

Figure 2. xGen Hyb Capture workflow. Targeted library fragments are separated from off-target fragments using hybridization capture. First, xGen Universal Blockers are mixed with prepared library fragments to prevent adapter-to-adapter hybridization. Blocked library fragments are then annealed to the 5′ biotinylated oligonucleotide probes from xGen Custom Hyb Capture Panels. The probe and fragment duplexes are then separated from the unbound fragments by streptavidin-coated magnetic bead purification. The resulting library is highly enriched for fragments of interest.

xGen Custom Hyb Capture Panel probe design for targeted methyl-seq

Designing hybridization capture panels for bisulfite-converted templates is uniquely challenging as the sequence of each strand of the input DNA duplex is no longer complementary following conversion. Unmethylated cytosine bases are converted to uracil during the bisulfite treatment and replicated as thymidine, whereas methylated cytosine bases will remain unchanged. Therefore, strand-specific probes are required at each target locus. Another consideration in probe design is that unmethylated strands will convert to high AT, and hypermethylated strands may be GC rich such that a broad GC range will be produced. To account for this and to ensure the complete methylation state is represented in captured libraries, each strand specific probe has two designs, assuming both a completely methylated and completely unmethylated status of each target locus, yielding in total 4 unique probes per target locus (Figure 3). Additionally, the xGen Hyb Capture workflow is well-suited to capture partially methylated versions of the target by allowing for some probe: target mismatched bases during hybridization. Utilizing the complete targeted methyl-seq workflow and probe design strategy provides adequate coverage and representation of all methylation states including fragments with intermediate methylation.

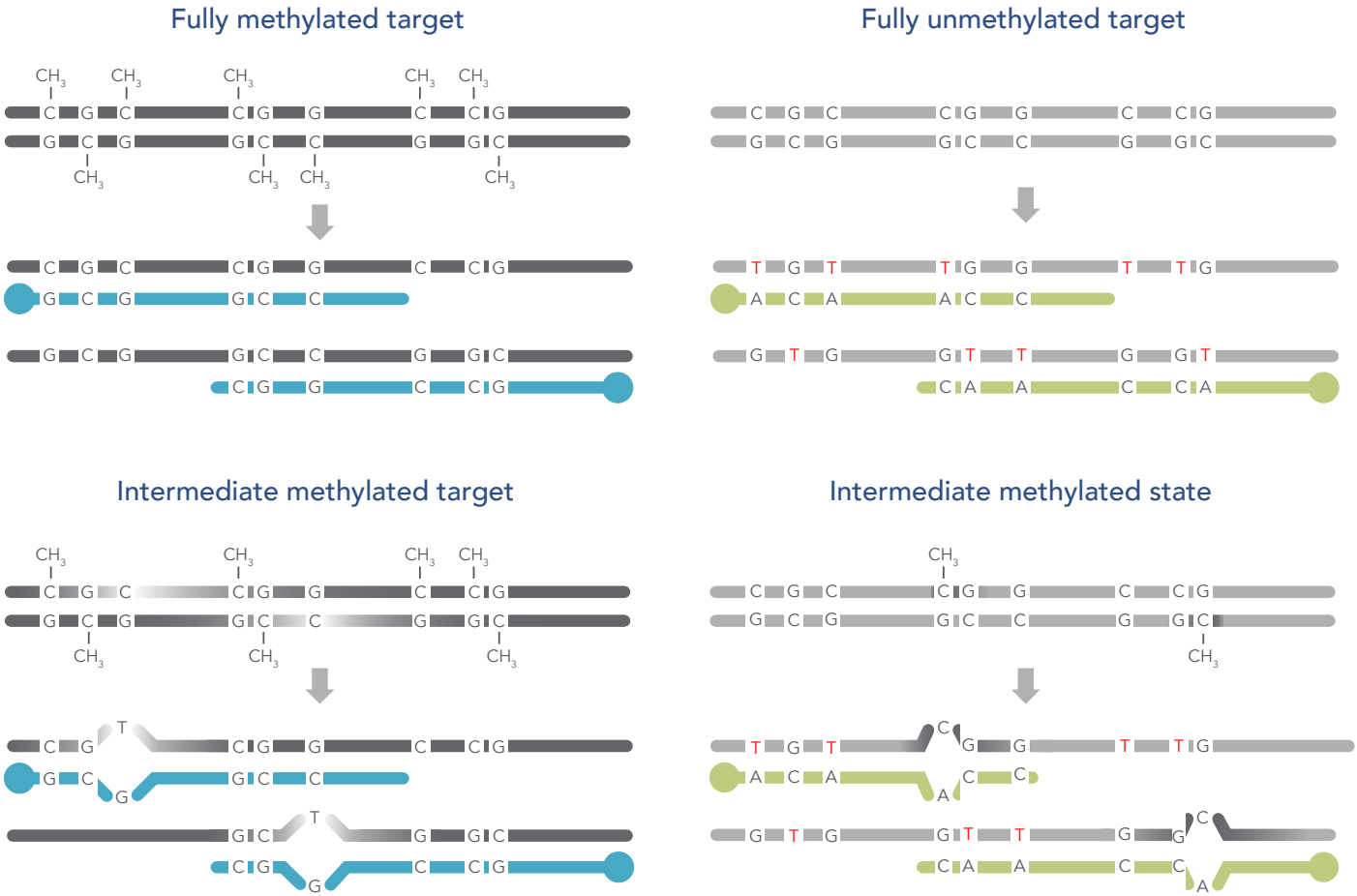

Figure 3. Schematic of methylation-specific xGen Custom Hyb Capture Panel probe design. Biotinylated probes in an xGen Custom Hyb Panel for targeted methyl-seq are designed to capture fully methylated, partially methylated, and fully unmethylated targets on both the top and bottom strands of each duplex. When targets are either fully methylated (blue probes) or fully unmethylated (green probes), the four probes hybridize to both strands of the target without any mismatches (top schematics). For partially methylated versions of the same target sequences, the xGen Hyb Capture workflow is designed to allow mismatches, and therefore, probes are able to capture the intermediate states (lower schematics).

Methods

Human methylated and non-methylated DNA standards (Zymo Research) from genomic DNA (gDNA) were combined in varying ratios to create a range of samples representing 0%, 15%, 50%, 85%, and 100% genome-wide methylation levels. Samples were sheared to 300 bp using a Covaris S220 and bisulfite conversion was performed in triplicate using the EZ DNA Methylation-Gold Kit (Zymo Research) starting with 50 ng from each of the five representative mixes. All bisulfite-treated material was carried forward to create xGen Methyl-Seq DNA libraries using modifications outlined in Appendix A of the xGen Methyl-Seq DNA Library protocol. A second set of libraries were prepared as described above using 5, 10, or 25 ng of cfDNA reference material (RM) (Horizon HD833) into bisulfite conversion and without any mechanical shearing. A methylation-specific xGen Custom Hyb Panel was designed to enrich for five genes associated with methylation biomarkers, covering a total region of 128 kb. Hybridization capture was performed on all libraries following the xGen Hybridization Capture protocol, hybridizing three libraries per capture for 16 hours. Sequencing was performed on a NextSeq™ (Illumina) platform using 2x75 bp reads with 10% phiX spike-in. For all analyses, 10 bp were trimmed from each end due to the low complexity tail added during the Adaptase step (see the application note, Tail trimming for better data, for more information). Each captured library was sub-sampled to 2 million total reads, and one WGBS from cfDNA RM was sub-sampled to 860 million total reads. Targeted sequencing data including alignment to the human genome (hg38) and methylation analyses were performed using Bismark (v0.22.3) and Picard (v2.18.9). For all samples, conversion efficiency was consistently ≥99% (data not shown).

Results

Hybridization capture results for gDNA libraries comparing a broad range of methylation levels

Key NGS metrics were used to assess the efficiency of the hybridization capture workflow. Enriched gDNA libraries showed high mapping rates of ≥83%, and high on-target rates of ≥84% across the entire range of simulated methylation levels (Figure 4A). Samples with lower methylation levels had less base composition complexity after conversion and can tend to be difficult to capture if probe specificity is affected. A consistent and high on-target rate was seen for all libraries suggesting this targeted methyl-seq workflow and capture panel design strategy can be used to determine methylation across a wide range of methylation levels without the loss of sequencing reads due to capture specificity. A high level of capture uniformity was seen for all enriched libraries across the entire range of simulated methylation levels demonstrated by consistently low fold-80 scores of ≤1.4 (Figure 4B), with minimal difference in the number of unique molecules seen in each library as shown by HS library size (Figure 4B). All captured libraries showed a comparable mean target coverage of >100X (Figure 4C), with strand bias values close to 1 indicating a similar number of reads from each original DNA strand were captured. Equivalent read depths and strand bias values close to 1 across all methylation mixes suggest little to no bias is introduced by strand-specific probe design during hybridization capture for the targets analyzed in this custom panel, regardless of the simulated methylation level.

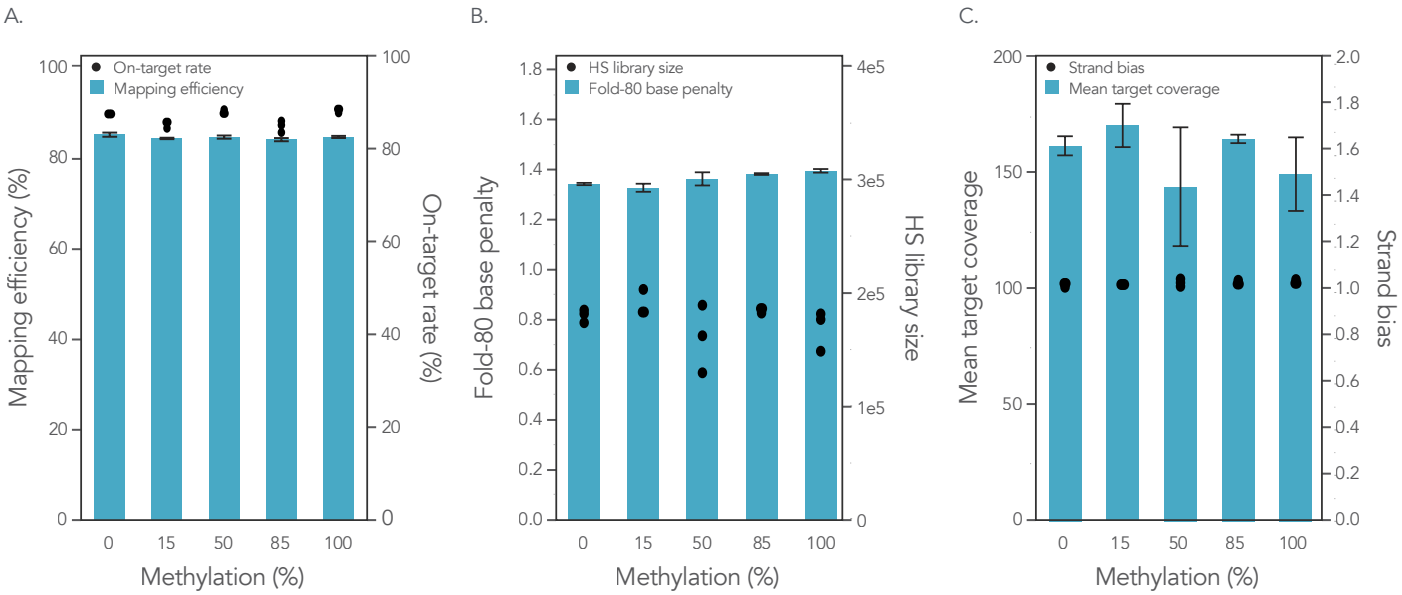

Figure 4. Target enrichment metrics for xGen Methyl-Seq libraries from gDNA methylation mixes. Human methylation standards of 0% and 100% methylation were combined, sheared and bisulfite-treated to create libraries simulating a range of methylation levels. All libraries were multiplexed (i.e., 3-plex) in an overnight hybridization capture using a 128 kb xGen Custom Hyb Panel. Sequencing was performed on a NextSeq™ (Illumina) instrument and subsampled to 2 million total reads/sample. The data shows (**A**) consistent high mapping efficiency and on-target rate, (**B**) equivalently low fold-80 scores with minimal differences in HS library size, and (**C**) consistent target coverage with no strand bias across all methylation mixes.

Quantification of methylation levels

The xGen Custom Hyb Panel was designed to capture post bisulfite-converted DNA. As methylation status and final sequence composition is not always known, each probe is designed to assume either fully methylated or fully unmethylated target sequence. Due to biological samples varying in methylation status, it is important to be able to accurately identify methylation levels between hypo- and hyper-methylation. The xGen Hyb Capture system is designed to efficiently and uniformly enrich across a range of hypo- to hyper- methylated targets. Libraries generated from samples containing 0%, 15%, 50% and 85%, and 100% methylation were used to determine whether the capture panel design can correctly identify methylation signatures across a broad range of simulated methylation levels.

The percentage of observed CpG methylation was compared to the percentage of expected CpG methylation of the mixes. An R2 value of 1.00 demonstrates consistency in the workflow without being adversely affected by the range of methylation ratios across the samples (Figure 5). The base-level view shows a scalable differentiation in observed CpG sites across methylation mixes. Across the methylation ratios, the same CpG sites were identified, and the observed methylation levels matched the expected methylation levels of the controls (Figure 6). Overall, this demonstrates that the xGen Custom Hyb Panel can distinguish methylation signatures across a complete range of simulated methylation ratios that also represent a broad range of GC base composition.

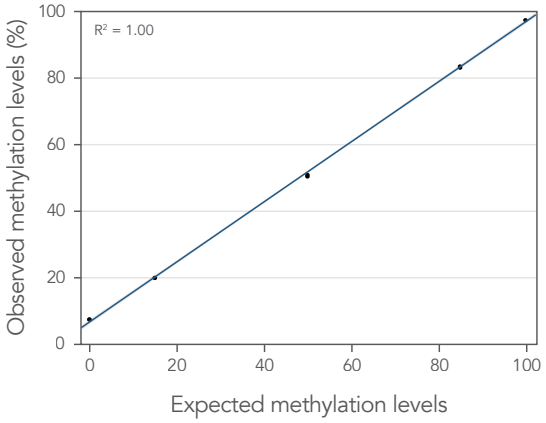

Figure 5. High correlation of CpG methylation levels across methylation mixes. Human methylation standards of 0% and 100% methylation were combined, sheared and bisulfite-treated to create libraries simulating a range of methylation levels. All libraries were multiplexed (i.e., 3-plex) in an overnight hybridization capture using a 128 kb xGen Custom Hyb Panel. Sequencing was performed on a NextSeq™ (Illumina) instrument and subsampled to 2 million total reads/sample. Methylation values were determined using Bismark and plotted against the expected methylation levels. A high correlation is seen between observed and expected methylation (R2 = 1).

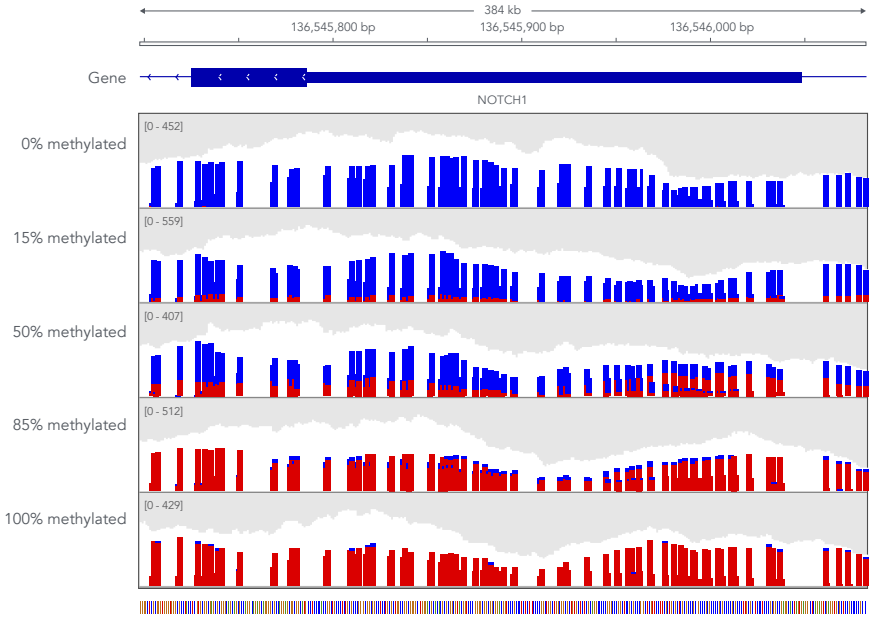

Figure 6. Integrated Genomics Viewer (IGV) image of NOTCH1 shows expected methylation across all levels following hybridization capture. Human methylation standards of 0% and 100% methylation were combined, sheared and bisulfite-treated to create libraries simulating a range of methylation levels. All libraries were multiplexed (i.e., 3-plex) in an overnight hybridization capture using a 128 kb xGen Custom Hyb Panel. Sequencing was performed on a NextSeq™ (Illumina) instrument and subsampled to 2 million total reads/sample. Each row represents one captured library from each of the methylation mixes starting with 0% in the top row through 100% at the bottom row. The blue color indicates an unmethylated base while the red color indicates a methylated base. The height of each bar corresponds to coverage of the site—with the scale of coverage shown at the upper left corner of each row. The number of sequencing reads identifying methylated or unmethylated bases scales with expected methylation values across the range of ratios.

Hybridization capture supports low input samples

As minimally invasive and longitudinal studies gain traction, cfDNA has become an increasingly utilized sample type for epigenetic research. However, cfDNA has relatively low abundance in the bloodstream which presents a challenge for research studies. The small amount of cfDNA that is extracted typically doesn't meet the minimum input requirements for commercial kits which cannot support <10 ng of DNA input. To assess the compatibility of low-input samples with the targeted methyl-seq workflow, libraries were prepared using 5, 10, or 25 ng of cfDNA reference material, which was followed by hybridization capture using the same xGen Custom Hyb Panel as above for target enrichment prior to performing NGS.

When subsampled to the same read level, key target enrichment metrics for low-input cfDNA RM libraries were comparable to those generated from the higher input gDNA libraries. The mapping rate for all samples was ≥84% with small standard deviations, and the on-target percentages for all inputs remained high at ≥84% for each sample (Figure 7A). This indicates that the targeted workflow and capture panel design strategy performed equivalently well on both low input amounts of cfDNA RM and 50 ng input amounts of gDNA. The custom xGen Hyb Capture Panel captured all expected methylation targets and the libraries showed equivalently low fold-80 scores as compared to the gDNA methylation mixes, which indicates a high capture uniformity for all sample inputs (Figure 7B). As expected, HS library size and coverage scaled with sample input amount while the 5 ng input showed the lowest number of unique molecules (Figure 7B and 7C). Again, a strand bias value of 1 is shown, which indicates that the xGen Custom Hyb Panel enriched both strands equally such that there was no capture bias of a specific strand (Figure 7C). To demonstrate the efficiency of the hyb capture workflow, the levels of cytosine methylation in CpG, CHG, and CHH (where H is A, C, or T) were quantified. The results show a similar percentage of C's methylated in CpG, CHG, and CHH across the range of sample inputs, which indicates an equivalently high sample recovery for all tested DNA input amounts and across all technical replicates (Figure 7D).

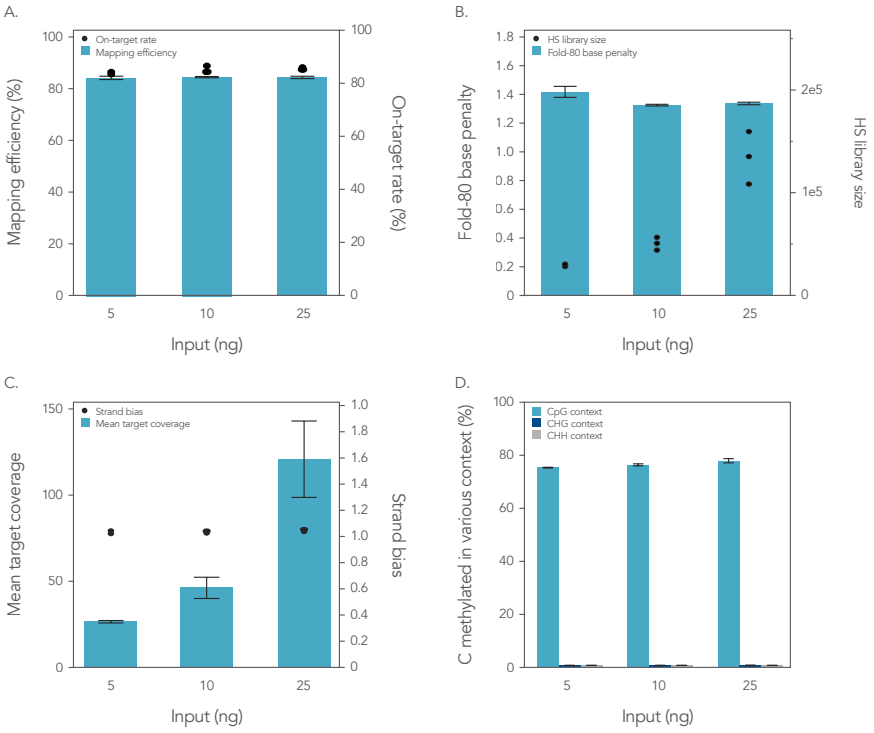

Figure 7. Target enrichment metrics for xGen Methyl-Seq libraries created from cfDNA reference material. Hybridization capture results using a 128 kb xGen Custom Hyb Panel with libraries made from 5 ng (N=2), 10 ng (N=3), and 25 ng (N=3) of cfDNA RM were sequenced on a NextSeq™ (Illumina) instrument and subsampled to 2 million reads per sample. The data showed equivalently (**A**) high on-target percentages and high mapping efficiency and (**B**) low fold-80 scores. As expected, HS library size (**B**) and (**C**) the mean coverage depth increased when higher amounts of input DNA were used. (**D**) A similar percentage of C's methylated in various contexts were seen across the target space for all sample inputs.

Concordant site-specific methylation between whole genome and target-enriched libraries from low input samples

To determine if target-enriched libraries preserved methylation signatures, the methylation percent per CpG site for a whole genome library and target-enriched libraries was calculated. The Pearson's correlation was determined by comparing these per-site percentages to a WGBS library across the shared target space (Figure 8). A high correlation ($r \geq 0.97$) of site-specific methylation was observed between WGBS and each of the target enriched libraries, as well as, for all the enriched library technical replicates ($r \geq 0.98$). This indicates that libraries which underwent hybridization capture reproducibly retained complete methylation information following target enrichment.

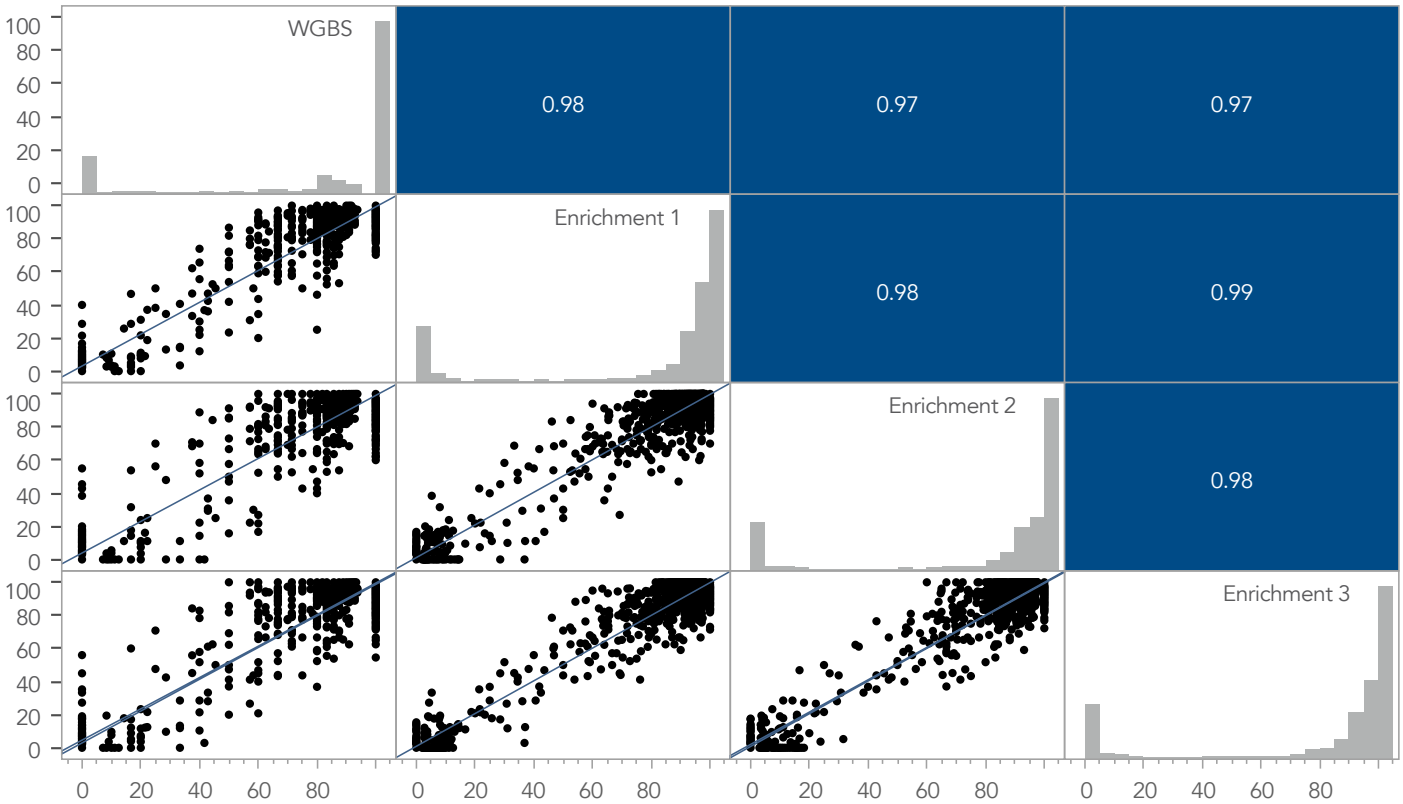

Figure 8. Pearson's correlation chart for target enriched xGen Methyl-Seq libraries vs. WGBS. Per-CpG site methylation for three captured libraries and one whole-genome library made from 10 ng cfDNA reference material. In total, 860 million reads for WGBS were used to compare the methylation percent per CpG site for any site that had ≥ 5X coverage in the shared target space. A high correlation of site-specific methylation is observed between whole-genome and target enriched libraries, including regions containing hypo-, hyper- and intermediate methylation.

Conclusions

The xGen Methyl-Seq DNA Library Prep Kit is a robust and flexible solution driven by Adaptase technology, a library preparation chemistry which directly tails and performs adapter ligation to single-stranded bisulfite treated DNA fragments. In traditional bisulfite sequencing, the library adapters are added prior to bisulfite conversion, which can result in fragmented non-functional library molecules. A unique single-strand strategy efficiently creates NGS libraries that allow researchers to measure methylation levels from a wide range of sample input amounts and over a wide range of methylation levels. Libraries enriched using a methyl-specific xGen Custom Hyb Panel show methylation signatures that are highly correlated with whole genome sequencing data from the same sample, and therefore, targeted methyl-seq provides a cost-effective way to assess methylation of key targets associated with different disease states. A unique xGen Custom Hyb Panel design targeting both DNA strands post-conversion resulted in efficient capture of a complete range of simulated methylation levels and allows for deeper sequencing of regions of interest in comparison to whole-genome sequencing. Together, the xGen Methyl-Seq Library Prep Kit and an xGen Custom Hyb Panel create a meaningful tool for accelerating focused epigenetic research.

References

1. Esteller M. **Epigenetics in cancer.** N Engl J Med. 2008;358(11):1148-1159.
2. Moarii M, Boeva V, Vert JP, Reyal F. **Changes in correlation between promoter methylation and gene expression in cancer.** BMC Genomics. 2015;16:873.
3. Diehl F, Schmidt K, Choti MA, et al. **Circulating mutant DNA to assess tumor dynamics.** Nat Med. 2008;14(9):985-990.
4. Gouil Q, Keniry A. **Latest techniques to study DNA methylation.** Essays Biochem. 2019; 63(6): 639–648.
5. Ziller MJ, Stamenova EK, Gu H, Gnirke A, Meissner A. **Targeted bisulfite sequencing of the dynamic DNA methylome.** Epigenetics Chromatin. 2016;9:55.
6. Tanaka K, Okamoto A. **Degradation of DNA by bisulfite treatment.** Bioorg Med Chem Lett. 2007;17(7):1912-1915.

Cost-effective targeted methyl-seq using an xGen™ Custom Hyb Capture Panel and the xGen Methyl-Seq DNA Library Prep Kit identifies DNA methylation in low input samples consistent with WGBS

Technical support: applicationsupport@idtdna.com

For more than 30 years, IDT's innovative tools and solutions for genomics applications have been driving advances that inspire scientists to dream big and achieve their next breakthroughs. IDT develops, manufactures, and markets nucleic acid products that support the life sciences industry in the areas of academic and commercial research, agriculture, medical diagnostics, and pharmaceutical development. We have a global reach with personalized customer service.

> SEE WHAT MORE WE CAN DO FOR **YOU** AT **WWW.IDTDNA.COM**.

For Research Use Only. Unless otherwise agreed to in writing, IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Purchaser is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations.

© 2022 Integrated DNA Technologies, Inc. All rights reserved. Trademarks contained herein are the property of Integrated DNA Technologies, Inc. or their respective owners. For specific trademark and licensing information, see www.idtdna.com/trademarks.
Doc ID: RUO21-0602_001 04/22

IDT INTEGRATED DNA TECHNOLOGIES

custom oligos • qPCR • next generation sequencing • RNAi • genes & gene fragments • CRISPR genome editing