Tunability and Optimization using the Twist Fast Hybridization Target Enrichment System for DNA Methylation Sequencing

Kristin Butcher, Leonardo Arbiza, Siyuan Chen, Richard Gantt, Brenton Graham, Sabina Gude, Rebecca Nugent, Christina Thompson, Ramsey Zeitoun

1. Abstract

Next generation sequencing (NGS) has given researchers the opportunity to link genetic factors to many disease states. However, the nucleotides encoding genomic DNA are only part of what influences cellular function and overall health. Other factors, including DNA methylation modifications on adenines and cytosines, can affect gene expression through epigenetic processes. Cytosine methylation is most commonly found at CG sequences in the genome, referred to as a CpG site, and is widely used to regulate gene expression in a cell-type specific manner. Targeted methylation capture panels can be used to focus sequencing efforts over areas containing genomic regions known to have variable and unevenness within methylation patterns.

In this study, we highlight the Twist Targeted Methylation Sequencing Product, in combination with several different sized custom designed DNA methylation panels. This product introduces a novel and innovative way to generate libraries used for methylation detection. We also show how this system can be tuned to best fit timelines and preferred final picard metrics while differentiating original methylation states We assess the Twist Fast Hybridization Target Enrichment systems within this product line, demonstrating how adjusting the fast wash buffer 1 temperature and the hybridization time can influence performance metrics. Finally, we establish an ideal protocol for the hybrid capture workflow with a fast wash buffer 1 temperature range that result in world class performance metrics and the ability to detect changes in DNA methylation modifications across various regions in the human genome. This includes performing the protocol while using an 8-plex, as well as using the Twist Methylation Enhancer to increase on-target sequencing.

2. Library Preparation for Methylation Detection

Identifying methylated cytosine residues is traditionally done through chemically converting unmethylated cytosines to uracils using bisulfite conversion. After PCR, unmethylated cytosines are seen as thymines, while methylated cytosines remain as cytosines (**Figure 2.1**) But, due to the harsh mechanical nature of this process, it often leads to DNA breaks that can make downstream processes difficult.

In the Twist Targeted Methylation Sequencing workflow, we use the NEBNext EM-seq kit developed by New England Biolabs. This process uses a reaction similar to bisulfite conversion via an enzymatic reaction featuring the APOBEC enzyme. Enzymatic conversion results in similar library prep yields as bisulfite conversion, and equal sensitivity of methylation detection at CpG sites (**Figure 2.2**) compared to bisulfite conversion. Enzymatic converted libraries show more uniform coverage in high GC regions (**Figure 2.3**), presumably due to the decrease in DNA breaks that would normally be prevalent using bisulfite conversion.

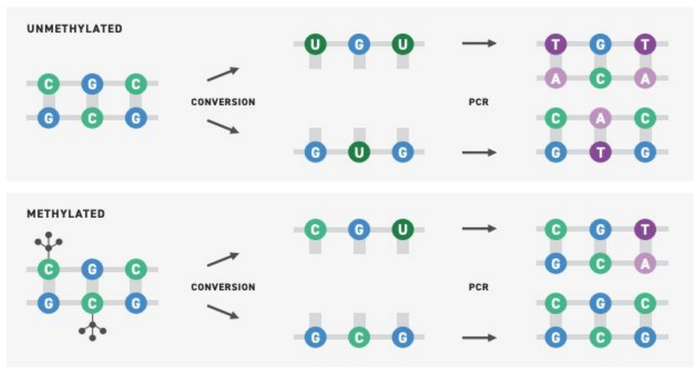


Figure 2.1: Methylation Conversion. Methylation sequencing involves enzymatic or chemical methods of converting unmethylated cytosines to uracil through deamination, while leaving methylated cytosines intact. During amplification, uracil is paired with adenine on the complementary strand, leading to the inclusion of thymine in the original position of the unmethylated cytosine. The end product is asymmetric, yielding two different double stranded DNA molecules after conversion (top row); the same process for methylated DNA leads to yet additional sets of sequences (bottom row).

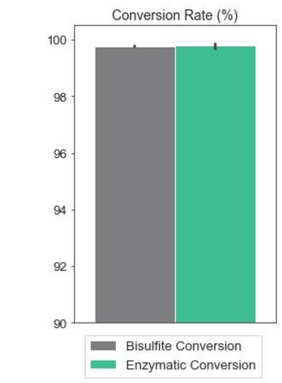


Figure 2.2: Conversion Rate Comparison for Bisulfite & Enzymatic Conversion. Conversion rates, measured as the percentage of cytosines converted to thymine in non-CpG sites, are >99.5% for both library conversion methods.

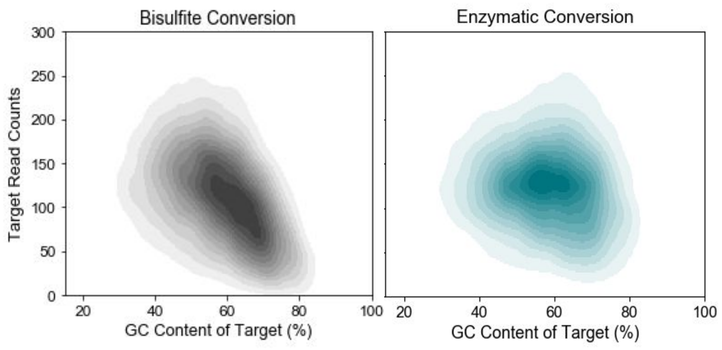


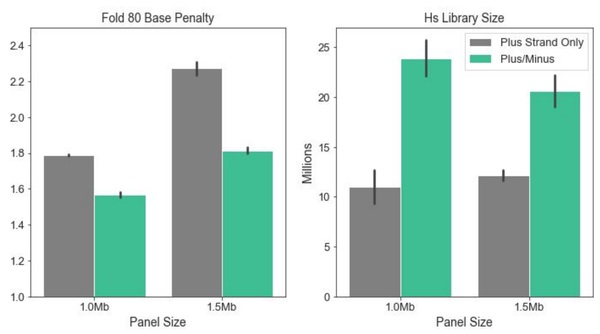
Figure 2.3: Coverage by Target GC Content for Bisulfite and Enzymatic Conversion. Both library conversion approaches are compatible with the Twist Fast Hybridization Target Enrichment System, although improved hybrid selection metrics are observed for libraries prepared with the enzymatic conversion approach. High GC target regions are associated with lower coverage when using the bisulfite conversion method (grey), while a less severe bias is observed when using the enzymatic conversion method (teal).

3. Customizable Panel Designs

Twist Bioscience combines precise oligonucleotide synthesis with a scalable silicon-based manufacturing platform to generate high-performing probe panels for hybrid capture. We complement this powerful technology with proprietary design algorithms and a rapid iteration pipeline to enable the quick design, synthesis, and optimization of custom panels. DNA inputs that are enzymatically or bisulfite converted prior to capture leads to added design intricacy as probes are designed to targets in each of the four possible converted sequences (**Figure 2.1**). The target genome grows several fold in content while non-methylated cytosines are converted to thymines leading to a reduction in overall sequence complexity.

Twist has developed a methylation-specific baseline design pipeline targeting the four possible converted sequences. This methylation-specific panel design pipeline allows for adaptive tunability capabilities by providing better control of off-target rates without adjustments to other important picard metrics. As highlighted in **Figure 3.1**, the Twist methylation design pipeline minimizes synthesis of unnecessary probes, decreasing off-target by eliminating the worst offending sequences from the final panel design. (**Figure 3.1**).

Our designs target the products derived from conversion of fully methylated and fully unmethylated DNA and target both the plus and minus strand. In **Figure 3.2** we demonstrate the importance of designing against both the plus and minus strand versus targeting only the positive strand. Uniformity, measured by Fold-80, and the number of unique molecules captured, measured by Hs Library Size, are both negatively impacted in the plus stand only design.



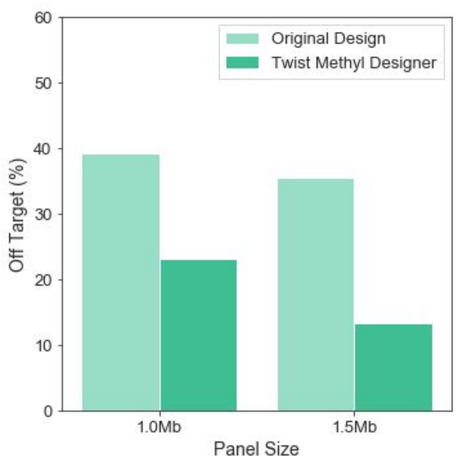


Figure 3.1: Reduction in Off-Target using the Twist Methylation Design Pipeline. The figure above shows reduction in off-target in 1.0Mb and 1.5Mb custom methylation panels generated through two design pipelines.

Figure 3.2: Improved Picard metrics with panels designed against both the plus and minus strands. The figure to the right show the Fold-80 (uniformity) and Hs Library Size (number of unique molecules) for two panels designed against the plus stand only (grey) or the plus and minus strands (Plus/Minus shown in teal).

Materials & Methods

Unless otherwise noted, sequencing was performed with a NextSeq® 500/550 High Output v2 kit to generate 2x151 paired end reads. Data was down-sampled to 250x of target size using aligned coverage, mapped using the Bismark Aligner, and analyzed using Picard Metrics with a mapping quality threshold of 20.

4. Twist Fast Hybridization Target Enrichment System

While the Twist Targeted Methylation Sequencing workflow uses alternative and novel technology to generate libraries, hybridization efficiency and optimization also play a part in this high throughput protocol. The Twist Fast Hybridization Target Enrichment System that is used within this workflow utilizes a faster hybridization time, arriving to hybridization equilibrium at 4 hours or less. This allows for the entire hybrid capture to be performed in less than one day, decreasing cost and turnaround time associated with hybrid capture. There are no detrimental metrics associated with this decrease in hybridization time moving from a 16-hour workflow to a 4-hour workflow.

The conversion of unmethylated cytosines to thymine in methylation detection library preparation leads to target molecules with reduced sequence complexity; a problem that has historically resulted in high off-target and poor uniformity in hybrid capture. In addition to our methylation panel design process, the buffer systems in our hybrid capture workflow have been developed to provide excellent sequencing performance in 4 hours or less. The increase in thymines that represent the unmethylated cytosines can lead to a decrease in binding temperature for the overall DNA structure during the hybridization process. Therefore, this hybrid capture system relies on full completion of hybridization at a temperature of 60°C, which can lead to incorrect and unnecessary binding of library molecules that may not contain the target region. Post hybridization, washes are implemented to help eliminate these wrongfully bound off-target strands by increasing the temperature during the fast wash buffer 1 step. The most optimal temperature will often be dictated by the target region itself and the size of the panel, but we found that a temperature range between 63°C and 66°C is a good starting point. By doing this, the stringency of the wash is increased, lowering the amount of off-target molecules that are still attached, which can also impact uniformity and coverage (**Figure 4.1**).

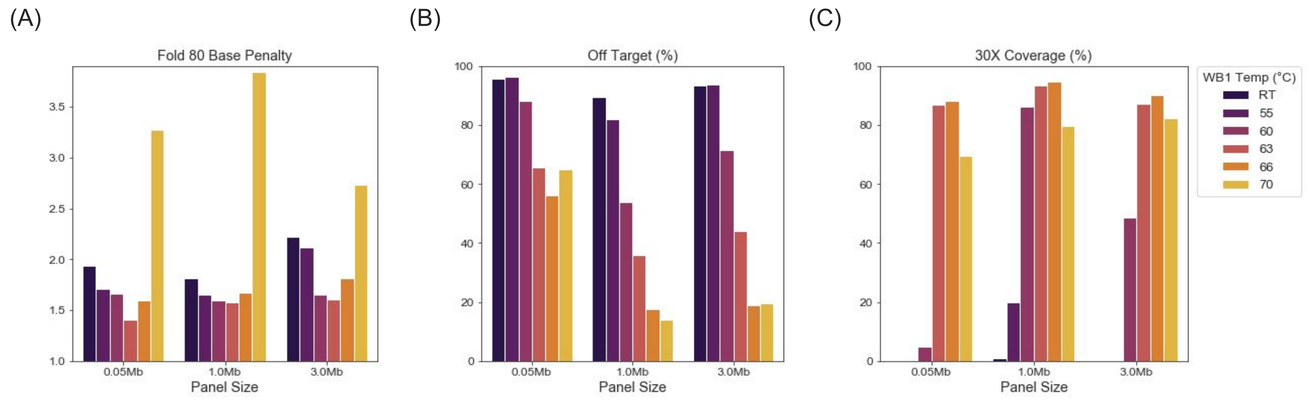


Figure 4.1: Picard Metrics using the Twist Fast Hybridization Target Enrichment System at Various Fast Wash Buffer 1 Temperatures. Hybrid capture was performed using different sized custom methylation panels and 200ng of library (NA12878; Coriell) and a 4-hour hybridization time with variation in the fast wash buffer 1 temperature. Custom methylation panels were designed using no stringency filters to best determine how off-target is impacted. A) The percentage of off-target molecules is plotted. B) Uniformity represented by the Fold-80 metric shows decreases as the fast wash buffer 1 temperature goes up, but starts to increase as temperatures higher than ~66°C. C) Percent target bases at 30x coverage initially increases as the fast wash buffer 1 temperature increases, but starts to decrease as the temperature increases over 66°C.

5. Methylation Enhancer

Commercial hybrid capture systems typically employ two types of blocker systems to reduce off-target capture for human samples: (1) Cot DNA, a product enriched for repetitive human gDNA and (2) adapter blockers, which are designed to limit adapter cross-hybridization. Due to the reduced sequence complexity generated by bisulfite or enzymatic conversion, we designed an additional blocker that prevents wrongfully targeted molecules from being pulled through the system.

We have designed a blocker that can be exclusively used in the Twist Targeted Methylation Sequencing workflow to decrease random off-target when using hybrid capture for methylation detection. This blocker, called Methylation Enhancer, reduces off-target in a variable fashion depending on custom panel target regions and methylation states of the genomic specimen. Despite the natural variability of target regions and genomic methylation states, there is no detriment to any hybrid selection metrics when using this blocker.

Methylation Enhancer is most helpful in reducing off-target when using a decreased hybridization time compared to the longest hybridization time of 4 hours. The flexibility of the Fast Hybridization Target Enrichment System gives researchers the opportunity to decrease hybridization time in a manner that best fits their schedule. When including Methylation Enhancer into the process, the hybridization reaction can be done in less time while still getting outstanding off-target metrics (**Figure 5.1**). By doing this, final data using a 2-hour hybridization time simulates the data produced after using a 4-hour hybridization time.

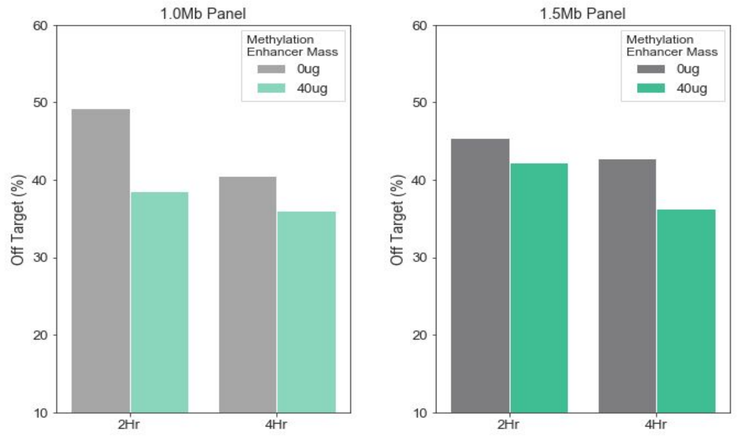


Figure 5.1: Performance of the Methylation Enhancer. Decreasing hybridization time in the Twist Fast Hybridization Target Enrichment system can often lead to higher off-target. When using Methylation Enhancer while decreasing hybridization time, the off-target metric is rescued. Hybrid captures were performed using a 1.0Mb and 1.5Mb custom methylation panel and 200ng of library (NA12878, Coriell) with a fast wash buffer 1 temperature of 63°C. Custom methylation panels were designed using no stringency filter to best determine how off-target is impacted. Grey bars represent reaction with no Methylation Enhancer, while the green bars represent reactions 40ug using Methylation Enhancer, each shade of green standing for a unique custom panel.

6. Multiplexing

High throughput and cost reduction for any experimental set-up continues to be a demand for all laboratories, regardless of area of study. This includes reducing protocol hand-on time, decreasing pipette-use, and lowering the amount of reactions used in a process. Using unique barcodes during the library generation protocol, or multiplexing, can help by lowering the amount of reactions needed during the hybridization process.

To characterize this process in the Twist Targeted Methylation Sequencing workflow, we performed 8-plex hybrid captures using a custom 1.0Mb methylation panel and compared it to single plex data. We performed these using the Methylation Enhancer to determine how the blocker works in a multiplex system. We also used a higher fast wash buffer 1 temperature compared to what we used in **Figure 5.1** in order to increase the wash stringency and decrease the off-target molecules brought through our system. By using the multiplex system, we observed a decrease in off-target, consistent uniformity represented by the Fold-80 metric, excellent coverage across thar target area at 30x, and an increase in unique molecules represented by the Hs Library Size metric (**Figure 6.1**). The multiplex system is compatible with this workflow.

Results demonstrate the unique capability of Twist's technology in target enrichment panels for methylation sequencing. We enable highly optimized designs out of the box with the flexibility of being able to choose among different levels of design stringency. The Twist Targeted Methylation Sequencing end-to-end workflow starts with genomic material and finishes with material to load onto any Illumina-based NGS platform, taking approximately a total of 18 hours. With decreases in protocol time, in part due to our novel library preparation method and to our efficient hybridization technology, researchers can go from genomic material to aligned sequencing data in less time than ever before with excellent hybrid capture performance metrics.

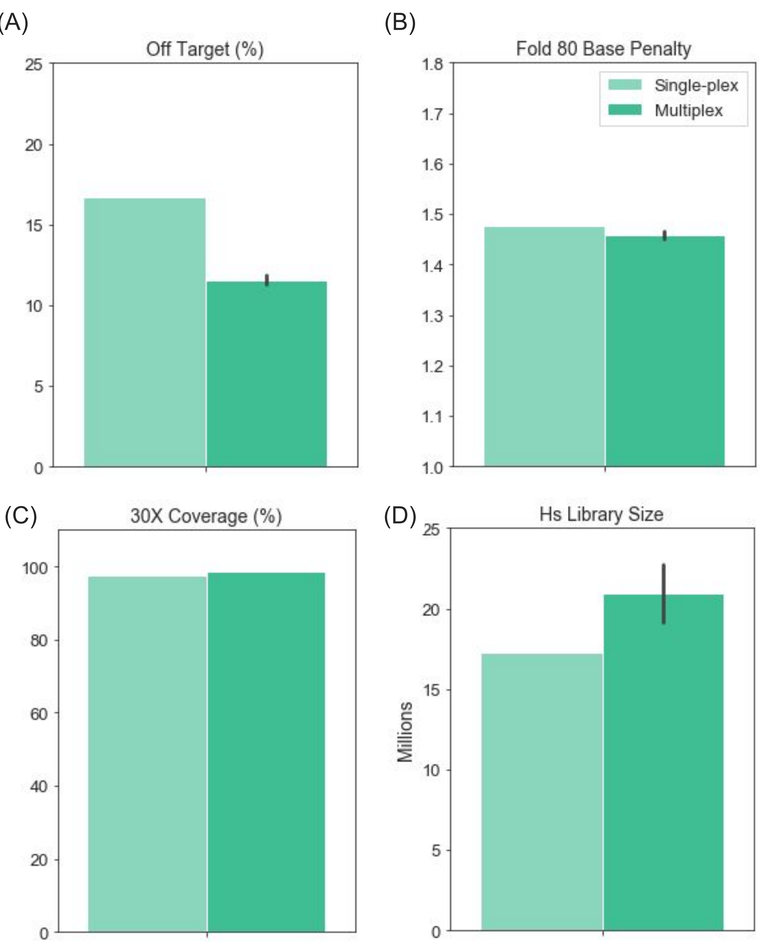


Figure 6.1: Picard Metrics using a Multiplex System. An 8-plex library reaction was pooled to generate a multiplex containing a total of 1500ng of library (NA12878, Coriell). Hybrid capture was performed using a 1.0Mb custom methylation panel and a 2-hour hybridization time at a 65°C fast wash buffer 1 temperature. The custom methylation panel was designed using no stringency filter to best determine how off-target is impacted. A) The percentage of off-target molecules go down when a multiplexed system is utilized. B) Uniformity represented by the Fold-80 metric shows a constant value. C) Coverage at 30x slightly increases while using a multiplex system. D) Hs Library Size improves in multiplex captures.

Financial Disclosures: Nothing to Disclose