Targeted Methylation Sequencing

INTRODUCTION

In the decades since sequencing the human genome, researchers have linked genetic factors to many disease states. However, the nucleotides encoding genomic DNA are only part of how genetics influence cellular functions and overall health; epigenetics (changes, in some cases heritable, that occur without modifying DNA sequences) also play a key role. DNA methylation is a well studied epigenetic marker chemically modifying cytosine and adenine. 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-specific manner.[1,2] Genome wide association studies (GWAS) link changes in DNA methylation to complex diseases like cancer and obesity and complex biological states like aging and development.[1,3–5] In recent years, methylation sequencing has expanded beyond gene expression and is starting to gain traction in disease diagnosis. For example, studies have shown differential methylation is a particularly informative and sensitive marker for cancer detection, regardless if methylation is associated with gene expression.[6]

Methylation Sequencing and Capture

Methylation sequencing involves enzymatic or chemical methods leading to the conversion of unmethylated cytosines to uracil through a series of events culminating in deamination, while leaving methylated cytosines intact (Figure 1). 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, Figure 1); the same process for methylated DNA leads to yet additional sets of sequences (bottom row, Figure 1).

Target enrichment can proceed by pre- or post-capture conversion. Post-capture conversion targets the original sample DNA to the left, while pre-capture conversion targets the four strands of converted sequences on the right (Figure 1). While post-capture conversion presents fewer challenges for probe design, it often requires large quantities of starting DNA material as PCR amplification does not preserve methylation patterns and cannot be performed before capture. Therefore, pre-capture conversion is often the method of choice for low-input, sensitive applications such as cell-free DNA.

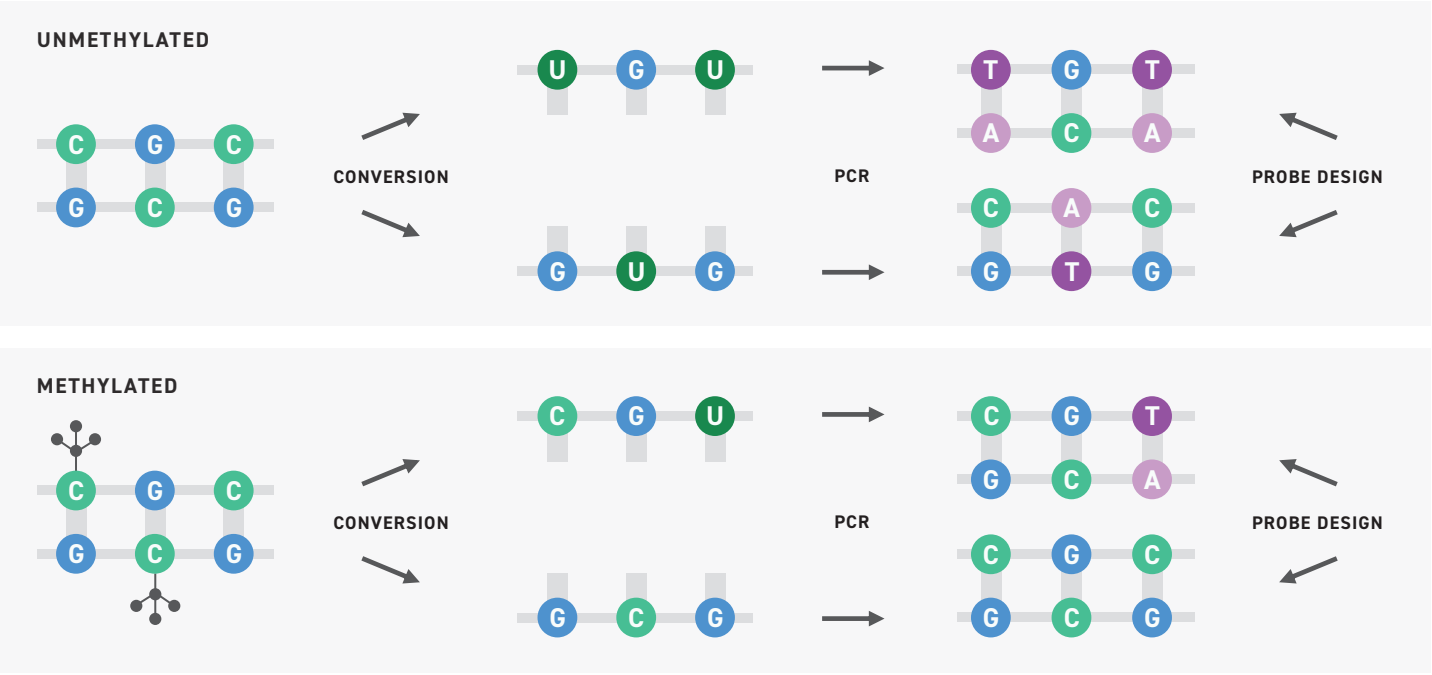

Figure 1. Methylation Conversion.

Panel Design

Pre-capture conversion leads to added intricacy as probes are designed to targets in each of the four possible converted sequences (Figure 1). Effectively, the target genome grows several fold in content while non-methylated cytosines are converted to thymine leading to a reduction in overall sequence complexity. This is the reason panels in this space have often been plagued with high off-target and reduced uniformity of coverage.

Here we sought to tackle the challenges of targeted methylation sequencing taking advantage of Twist's high quality oligo synthesis, capture solutions, and the data-driven methods that have enabled very high uniformity panels. Using Twist's newly developed methylation-specific baseline design algorithm and targeted methylation sequencing protocol, we demonstrate high performance metrics for panels of varying sizes. Using the data and responses learned, we implement optimized default design capabilities with different levels of stringency. While arrays are still widely used in methylation detection, high quality target enrichment panels provide an attractive alternative to static array designs for exploring dynamic and cell-specific methylation targets or poorly understood targets found in more elusive non coding regions at single base pair resolution.

MATERIALS AND METHODS

Genomic DNA samples from NA12878 (Coriell Institute) and EpiScope® hypo- and hypermethylated gDNA controls (<5% and >95% methylated HCT116 DKO gDNA, respectively) were mechanically sheared to a size of ~300 bp (on Covaris® ME220). Samples of various simulated methylation levels were prepared by blending sheared hypo- and hypermethylated controls. 500 ng of gDNA input were put into the Swift Accel-NGS® Methyl-seq DNA Library Kit in combination with bisulfite treatment (Zymo EZ DNA Methylation-Lightning Kit, Omega Bio-Tek Mag-Bind RxnPure Plus SPRI Beads, and KAPA HiFi Uracil+ DNA Polymerase). 200 ng of gDNA input were put into the NEBNext® Enzymatic Methyl-seq Kit. Sheared samples and libraries were verified with the Agilent BioAnalyzer 7500 and the Invitrogen Qubit Broad Range Kits.

Twist's four hour fast hybridization protocol was re-engineered for use with Twist methylation panels. Four methylation panels covering a range of different target sizes (0.05, 1.0, 1.5, and 3.0 Mb) were synthesized and used to evaluate performance. 200 ng of library was used for each single-plex capture, followed by 2x151 bp sequencing on an Illumina NextSeq 550 with a v2.5 High Output Kit. Alignment and methylation analyses were performed using Bismark 19.1 and Picard HsMetrics after sampling to a raw coverage of 250x per sample.

RESULTS

Testing a Variety of Different Panels and Sizes

While pre-capture conversion can enable highly sensitive epigenetic applications, key challenges originate from the reduced complexity of the genome after conversion. Compared to non-methylated panels, this generally leads to markedly high off-target (levels >50–60%), lower target coverage, and a strong reduction of capture uniformity (fold 80 base penalty values >2.5). Twist's targeted methylation sequencing protocol was tested using four panels of different size covering a wide range of different methylation targets. Using Twist's baseline panel design capabilities, the panels used in the following experiments showed off-target levels below the current industry standard with off-target values as low as 27%. The 0.05 Mb panel showed higher off-target compared to the other three panels. This is likely due to the nature of extremely small target size. Capture uniformity was also significantly better than industry standards of >2.5 fold 80 and reached values as low as 1.75 and 1.5, which are remarkable for this particular application. The duplication rate was very low among all 4 tested panels, indicating the capture step was efficient and able to retain high sample complexity throughout the workflow. Overall, with 250x raw sequencing, we achieved a coverage higher than 70% at 30x, even for the smallest panel.

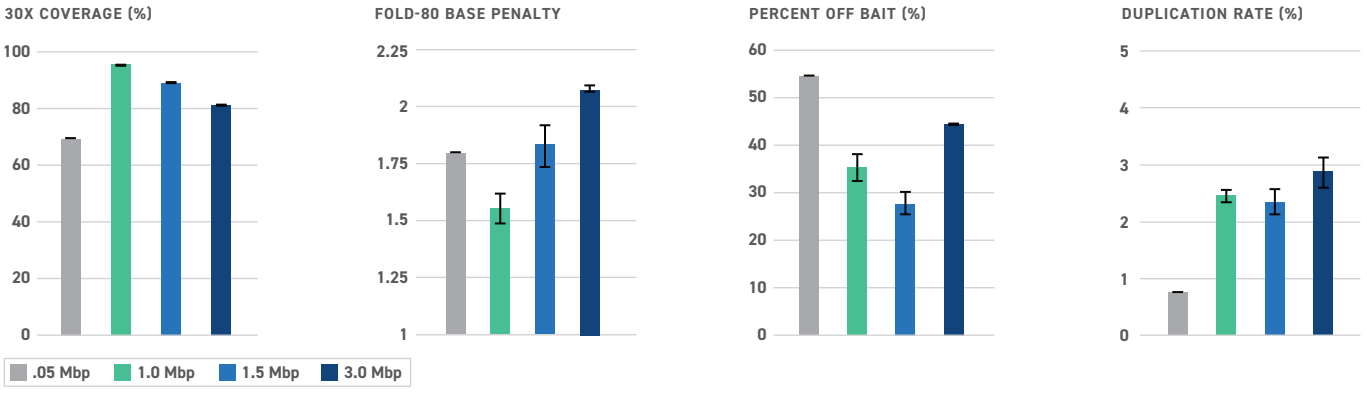

Figure 2. Panel Size.

Stringency Control for High Quality Pre-optimized Designs

Twist's adaptive panel design optimization algorithms enable the use of empirical data from capture experiments to learn about specific probe characteristics to quantitatively tune performance. This powerful method becomes particularly useful for methylation panels where controlling off-target rates becomes a priority. In addition, using data collected for over ~30,000 methylation targets, we derived informative sequence features and used these to develop optimized default panel design capabilities for custom panels with three levels of stringency. A 1 Mb panel was used as a demonstration of default panel designs with low, medium, and high stringency which provide increasing control of off-target rates while leading to only minor changes in other key metrics (Figure 3).

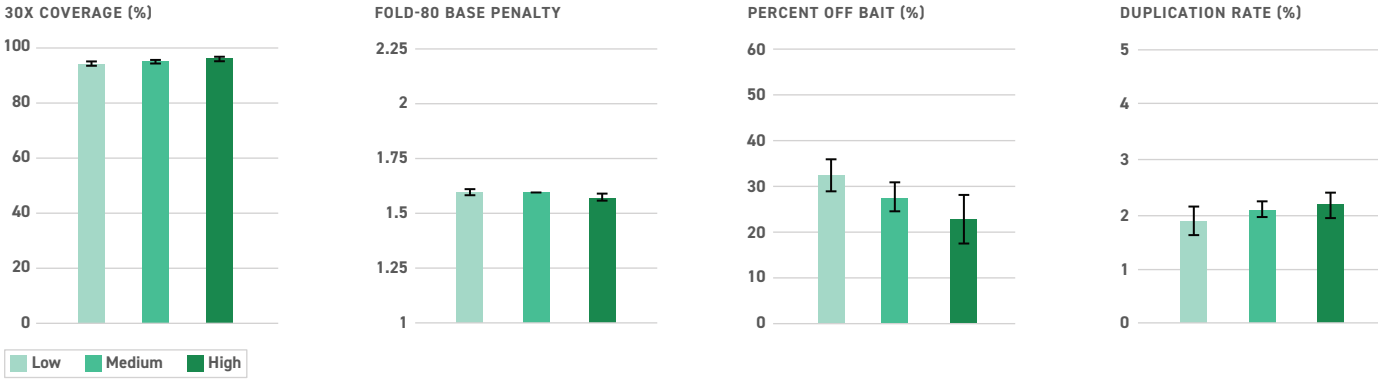

Figure 3. Design Stringency.

Capture Performance is Maintained Across Possible Ranges of Hypo- and Hypermethylation of Targets

To evaluate compatibility across a range of possible methylation levels, we performed captures on the medium stringency 1 Mb panel with gDNA libraries generated from hypomethylated and hypermethylated cell lines blended to final ratios of 0, 25, 50, 75, and 100% methylation, respectively. The figure below highlights key capture metrics with bars showing capture performance between differentially methylated samples. Metrics show little to no response to varying methylation levels, demonstrating the compatibility of the system with a wide range of methylation states including hypo- and hypermethylated DNA.

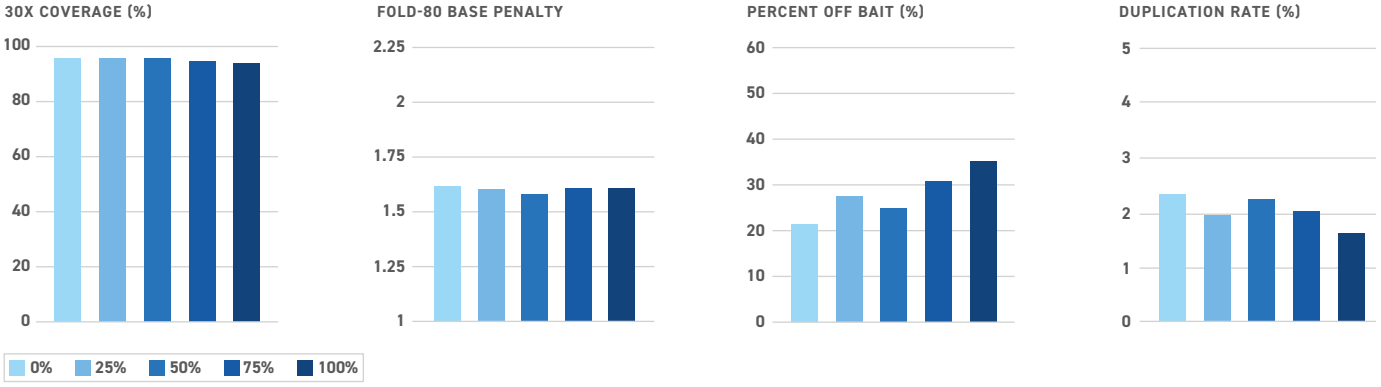

Figure 4. % Methylation Library.

Detection of Methylation Levels

Changes in methylation levels of promoters and other regulatory elements are emerging as some of the most sensitive markers available for the early detection of cancer.[7] We demonstrate targeted methylation sequencing can detect and quantify differential levels of DNA methylation. Hypo- and hypermethylated DNA were blended to different ratios and used for capture with the 1 Mb panel. Figure 5 highlights the detection of different DNA methylation levels along targets and individual CpG sites in the clinically relevant Cyclin D2 locus, which is known to change methylation states in certain cancers, e.g. breast cancer.[8]

Figure 5. Sensitivity of Methylation Detection.

Compatibility with Enzymatic and Bisulfite Based Assays

Detecting methylated cytosines involves the conversion of unmethylated cytosines to thymine while methylated cytosines are protected from conversion. Traditionally conversion occurred through a chemical bisulfite method. Other methods including enzymatic conversion of unmethylated cytosines (as used in results presented thus far) are being adopted in the field at increasing rates. Each conversion method has advantages and disadvantages, such as greater potential sensitivity of the enzyme to conversion reaction conditions or the context biased degradation of DNA by bisulfite.

Twist's targeted methylation sequencing is compatible with both enzymatic and bisulfite based approaches (Figure 6). Conversion rates, measured as the fraction of cytosines converted in non-CpG sites were >99.5% for both methods (data not shown). Overall capture metrics were comparably on the same order for both library preparation methods, though certain metrics such as uniformity, and off-target were reduced for the bisulfite method. We have evidence to show the reduced uniformity is at least partially due to the inherent GC bias introduced by the bisulfite based library preparation method (data not shown).

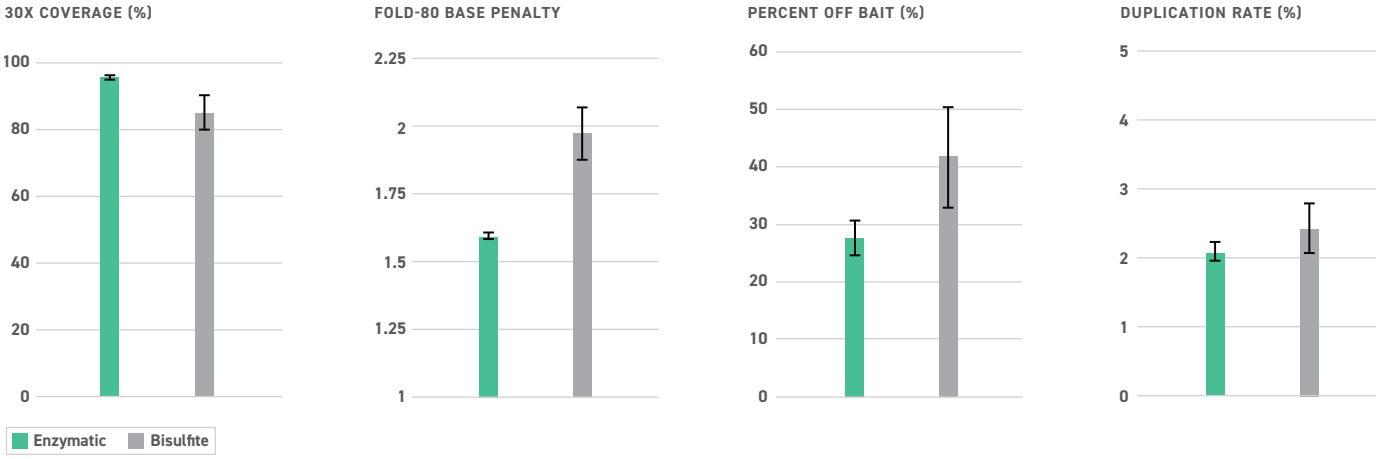

Figure 6. Methylation Conversion Method in Library Preparation.

CONCLUSION

Results demonstrate the unique capability of Twist's technology in target enrichment panels for methylation sequencing. We highlight excellent performance across a wide range of methylation targets and levels and enable highly optimized designs out of the box with the flexibility of being able to choose among different levels of design stringency. Compatible with enzymatic or bisulfite conversion, Twist's high uniformity methylation capture panels are new tools to push the envelope in the development of highly sensitive epigenetic applications.

REFERENCES

1. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13, 484–492 (2012).
2. Yang, X., Shao, X., Gao, L. & Zhang, S. Comparative DNA methylation analysis to decipher common and cell type-specific patterns among multiple cell types. Brief Funct Genomics 15, 399–407 (2016).
3. Loh, M., Zhou, L., Ng, H. K. & Chambers, J. C. Epigenetic disturbances in obesity and diabetes: Epidemiological and functional insights. Mol Metab 27S, S33–S41 (2019).
4. Ferrucci, L. et al. Measuring biological aging in humans: A quest. Aging Cell 19, e13080 (2020).
5. Pérez, R. F. et al. Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes. J Transl Med 17, 15 (2019).
6. Locke, W. J. et al. DNA Methylation Cancer Biomarkers: Translation to the Clinic. Front Genet 10, (2019).
7. Hao, X. et al. DNA methylation markers for diagnosis and prognosis of common cancers. PNAS 114, 7414–7419 (2017).
8. Evron, E. et al. Loss of cyclin D2 expression in the majority of breast cancers is associated with promoter hypermethylation. Cancer Res. 61, 2782–2787 (2001).

For research use only.