TECH NOTE

Highly Sensitive Methylation Detection Using Enzymatic Methyl-seq and Twist Target Enrichment

INTRODUCTION

DNA methylation plays an important role in cell growth by regulating gene expression. In some cases, aberrant cytosine methylation can lead to carcinogenic changes in gene expression. Historically, these modifications have been hard to identify in patient samples. Although next-generation sequencing (NGS) of DNA treated with sodium bisulfite has provided an efficient means for the detection of methylated cytosines, the chemical conversion process is damaging, leading to GC biased coverage and poor complexity[1]. Such protocols are therefore limited in their sensitivity.

The Twist Targeted Methylation Workflow introduces a complete solution that produces highly complex and uniform libraries for methylome analysis in a variety of settings, including novel ones like cancer screening by liquid biopsy. The end-to-end protocol achieves this by combining an innovative, enzymatic conversion process with a highly compatible and quick hybrid capture process using Twist Custom Methylation Panels. This technical note highlights the benefits of this workflow over existing protocols, discusses the use of controls throughout the workflow, and demonstrates the effects of panel target size and genomic DNA methylation level on downstream sequencing metrics.

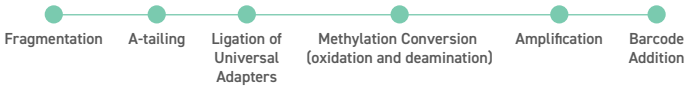
RESULTS

Library Preparation for Methylation Detection

Conventional protocols identify methylated cytosine residues by chemically converting unmethylated cytosines to uracils with bisulfite treatment. After PCR, unmethylated cytosines are sequenced as thymines, and methylated cytosines are sequenced as cytosines (Figure 1).

However, bisulfite treatment often leads to DNA breaks that can complicate downstream sample preparation and, ultimately, methylation detection[2]. Twist Bioscience® partnered with New England Biolabs® to offer NEBNext® Enzymatic Methyl-seq (EM-seq™) as part of the Twist Targeted Methylation Sequencing Workflow. This innovative process accomplishes the same conversion results as bisulfite treatment without the harshness of chemical conversion, yielding a superior end result.

Library preparation for methylation detection consists of 6 key steps:



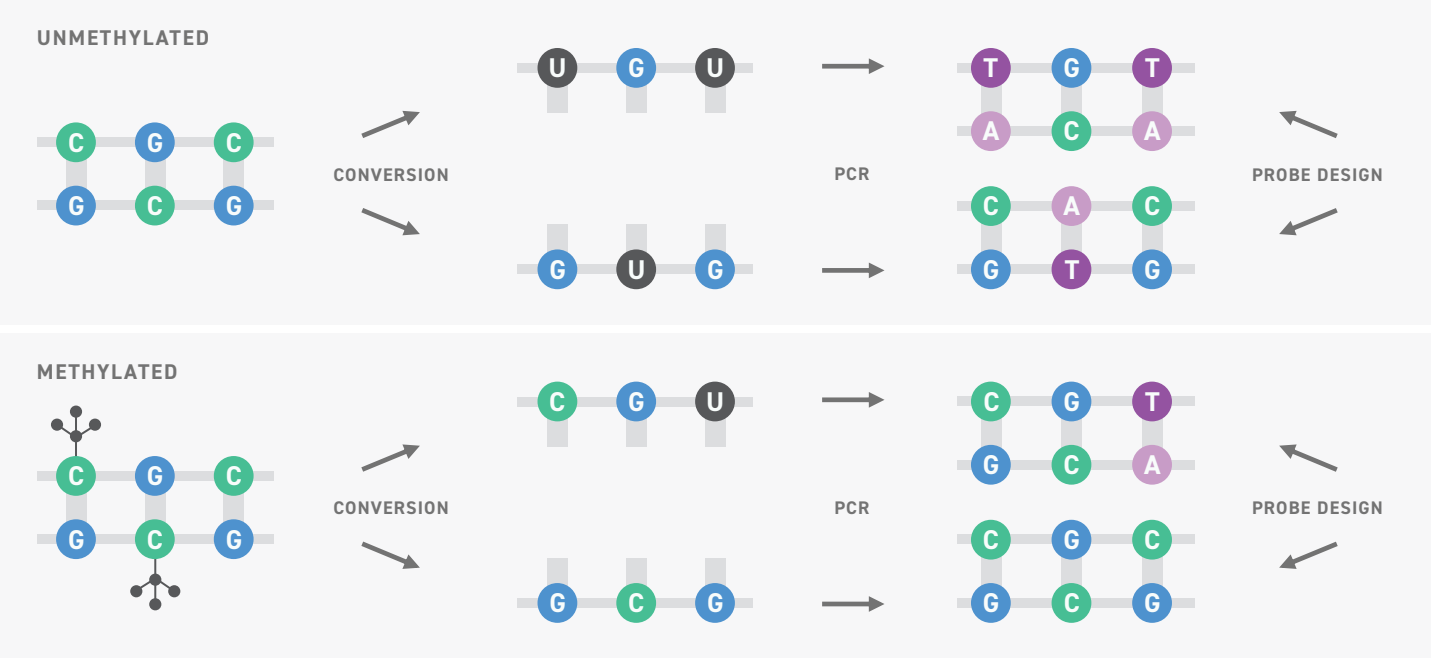


Figure 1. Methylation Conversion. Methylation sequencing involves enzymatic or chemical methods of converting unmethylated cytosines to uracil through deamination, while leaving methylated cytosines intact. During amplification, uracil is paired with adenine on the complementary strand, leading to the 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).

PUBLISHED: February 2021

The enzymatic and bisulfite methods convert unmethylated cytosines to thymines with similar efficiency (Figure 2). However, enzymatically converted libraries show increased and more uniform coverage in high GC regions (Figure 3), this is potentially a result of fewer DNA breaks at these sites.

The Twist Methylation System offers flexibility in fragment size by utilizing mechanical shearing before library preparation and enzymatic conversion. When using this library preparation method for methylation detection, Twist Bioscience recommends shearing genomic material to approximately 200 to 300 bp to generate final fragment sizes ranging between 350 bp and 450 bp, although the actual fragment size will depend on the initial integrity of input material. Library yield depends on the amount of input material and the number of PCR amplification cycles used during library preparation. Twist Bioscience recommends starting with 200 ng input material and nine amplification cycles. These recommendations ensure enough material is available for subsequent hybrid capture. The final quality control step confirms the fragment length distribution (Figure 4).

Use of the Control DNAs in the Twist Targeted Methylation Sequencing Workflow

Incomplete conversion increases the false positive rate of the assay, as unconverted, unmethylated cytosines are interpreted erroneously as methylated. To mitigate this, the conversion rate can be confirmed using DNA controls of known methylation levels. As part of the Twist Targeted Methylation Sequencing Workflow, the NEB EM-seq Methylation Library Preparation Kit includes CpG Methylated pUC19 DNA and Unmethylated Lambda DNA as optional controls. Both controls possess known levels of methylation, enabling an accurate determination of the conversion rate post-sequencing. Because they may lack complementary probes in target enrichment panels, these controls should not be subjected to hybrid capture. Instead, they should be stored until after hybrid capture and subsequently pooled with samples for sequencing.

To demonstrate the use of these controls in the Twist workflow, libraries were generated using the protocol described in Appendix A of the NEBNext EM-seq Methylation Library Preparation Protocol. Forty-eight microliters of each DNA control were combined together in a single reaction, and the mix was dried down using a speed vacuum concentrator. The resulting dried DNA was resuspended in 50 µl of 0.1X TE pH 8.0 and moved through the library process.

Table 1 shows the measured versus expected conversion efficiency and post-sequencing methylation level. EM-seq met the expected efficiency at higher than 99.5% conversion for both controls. The expected CpG methylation levels of the Unmethylated Lambda DNA and CpG Methylated pUC19 DNA controls are 0.5% and 95–98%, respectively. The measured CpG methylation levels for both controls matched the expected levels. This was calculated using the methylation caller used for analysis. These data indicate that DNA controls of known methylation levels can be used to ensure that the conversion process is complete and the assay's false positive minimized.

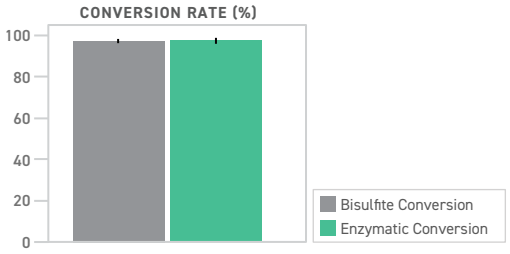


Figure 2. Conversion Rate Comparison for Bisulfite & Enzymatic Conversion. Both library conversion methods achieve conversion rates of >99.5%, as measured by the percentage of cytosines converted to thymine in non-CpG sites.

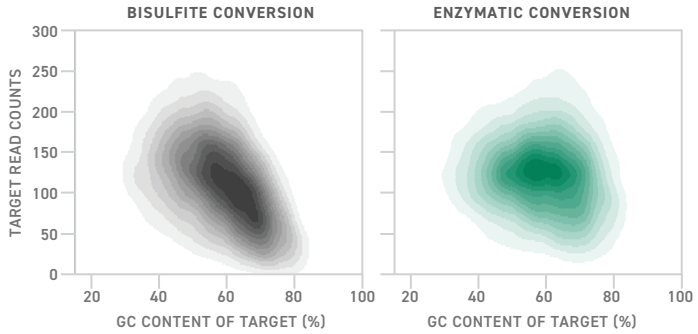


Figure 3. Coverage by Target GC Content for Bisulfite and Enzymatic Conversion. Both library conversion approaches are compatible with the Twist Fast Hybridization Target Enrichment System, although improved hybrid selection metrics are observed for libraries prepared with the enzymatic conversion approach. 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 (green).

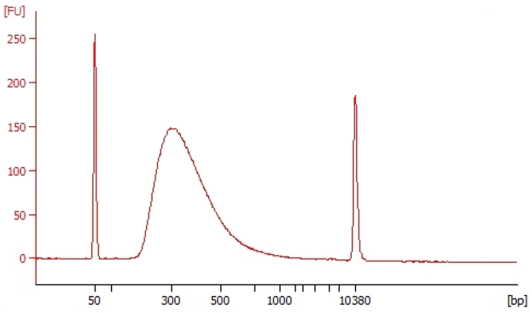


Figure 4. Final Fragment Analysis Trace. Final quality control step using the EM-seq conversion method for library preparation. The average peak length is approximately 375 bp.

METRIC	UNMETHYLATED LAMBDA DNA	CpG METHYLATED pUC19 DNA
Expected Conversion Efficiency	>=99.5%	>=99.5%
Measured Conversion Efficiency	99.77%	99.57%
Expected CpG Methylation Level	Up to 0.5%	95-98%
Measured CpG Methylation Level	0.22228%	95.7572%

Table 1. Expected versus Measured Conversion Efficiency and CpG Methylation Levels. Conversion efficiency and CpG methylation level results when converting the CpG Methylated pUC19 DNA and Unmethylated Lambda DNA controls with EM-seq.

Twist recommends using these controls when unsure about the enzymatic conversion efficiency during the EM-seq workflow. For every 96-reaction kit, there is enough of each control to make two separate control libraries.

Twist Target Enrichment for Methylation Detection

The Twist Targeted Methylation Sequencing Workflow also uses the Twist Fast Hybridization Target Enrichment System to achieve highly sensitive and customized methylation detection. This system reaches hybridization equilibrium without adversely impacting sequencing metrics despite its use of a faster hybridization time than the previous Twist workflow, arriving at hybridization equilibrium with no detriment in metrics despite decreased hybridization time.

The Twist Fast Hybridization Target Enrichment System involves 5 major steps:

This system enables multiplexing (up to 8-plex) for higher throughput and reduced hands-on time and pipette use. The Twist Targeted Methylation Sequencing Protocol recommends best practices that are compatible with a variety of custom methylation panels. However, as with all target capture systems, multiple factors can influence the final capture data quality, including the size and methylation level of the target region. The impact of these factors on downstream sequencing metrics is discussed below.

Target Region Size

Many factors related to custom target regions influence the final targeted sequencing metrics; optimization may be needed for best performance. For example, very small panel designs (<0.5 Mb) or panels with high GC content in the target region, are particularly sensitive to small changes in panel design. The optimal trade-off between inclusiveness and off-target control depends on characteristics of the target region and the panel's intended application. During the panel design process for example, a researcher working with a medium sized panel and a low number of samples may prefer to keep certain probes, even if they require additional sequencing to balance increased off-target capture. By contrast, those working with a much smaller panel (where off-target capture increases the required sequencing relative to rest of the panel more quickly) or with very large numbers of samples (where modest increases in cost can quickly add up), may prefer to use more stringent design conditions to optimize cost.

To evaluate the relationship between panel size and sequencing metrics, three different panels were used with the Twist Targeted Methylation Sequencing Workflow. Together, the panels spanned a wide range of methylation targets and panel sizes: 0.5 Mb, 3 Mb, and 50 Mb (Figure 5). The largest panel used in this study gave off-target levels close to 7%, with all panels registering off-target levels under 10%. Capture uniformity (fold-80 base penalty) was exceptional for all target sizes, reaching values between 1.4 and 1.7. The proportion of probes with 30X coverage was higher than 90% for all panels. However, the increase in duplication rate with panel size warrants consideration when defining what key metrics to prioritize.

Identification of differing Target Region Methylation Level

Methylation levels differ across the genome. Because differential methylation levels can be used for early detection of specific cancers, it is vitally important that the protocols used to detect methylation are highly compatible with custom panel designs and are capable of identifying hyper- and hypomethylated regions. Conversion leads to a decrease in sequence complexity, which can cause issues downstream in the hybrid capture step. However, these issues can be mitigated with the right combination of library preparation reagents, hybrid capture reagents, and custom panel design, resulting in probe coverage that is evenly distributed across regions with varying AT/GC content and methylation levels. Twist Custom Panels for targeted methyl-seq are built with 4 probe species per target to optimally capture the methylated and unmethylated forms of both the sense and antisense DNA after conversion (See Figure 1, right side).

To demonstrate the evenness of coverage that can be obtained with the Twist Targeted Methylation Sequencing Workflow, probe coverage plots were generated using a 1.5 Mb custom panel with

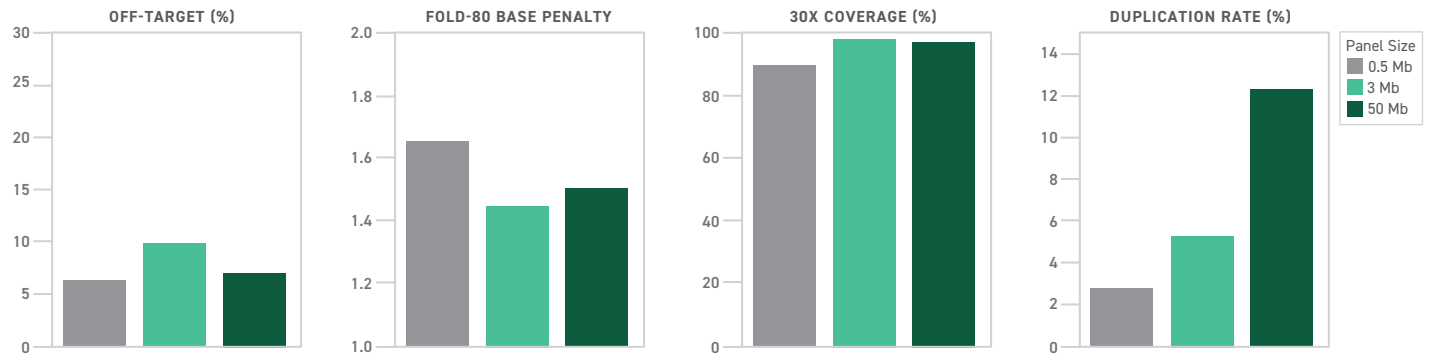


Figure 5. Size of Target Region within Custom Panel and its relationship to Picard Metrics. Key capture metrics for methylation panels covering target sizes of 0.5 Mb, 3 Mb, and 50 Mb using the Twist Targeted Methylation Sequencing Protocol and a single-plex reaction. Twist recommended capture conditions, including 2 µl of Methylation Enhancer, a Wash Buffer 1 temperature of 65°C, and a 2-hour hybridization time were used in each reaction. Sequencing was performed with a NextSeq® 500/550 High Output v2 kit to generate 2x76 paired end reads. Data was down-sampled to 200x aligned coverage relative to the panel target size, mapped using the Bismark Aligner, and analyzed using Picard Metrics with a mapping quality threshold of 20.

hyper- and hypomethylated genomic DNA input material using two different conversion systems: EM-seq and bisulfite treatment. Figure 6 shows an even distribution of target read counts for both methylation levels using the EM-seq conversion method (teal). By contrast, an industry-leading bisulfite conversion process (grey) resulted in comparatively uneven target read counts. Thus, Twist custom panel design efficiently captures all targeted DNA species —regardless of target methylation level—when used with the Twist Targeted Methylation Sequencing Workflow.

To look more closely at the performance of Twist Targeted Methylation Sequencing Workflow across a wide range of methylation levels, libraries of varying methylation levels were generated and captured using a 1 Mb panel. These libraries were produced by mixing hypo- and hypermethylated control human cell lines (EpiScope® Unmethylated HCT116 DKO gDNA PN# 3521 and EpiScope Methylated HCT116 gDNA PN# 3522, Takara Bio USA) to create libraries with 0, 25, 50, 75, and 100% methylation. Figure 7, which shows key capture metrics, highlights the lack of relationship between methylation level and final Picard metrics. These data demonstrate the compatibility of the Twist Methylation System with both hypo- and hypermethylated DNA regions.

CONCLUSION

Twist Bioscience now provides an end-to-end, targeted methylation detection workflow that combines an enzymatic conversion from NEB, with custom methylation panels and optimized target enrichment reagents. Overall, the Twist Targeted Methylation Sequencing Workflow enables highly sensitive methylation detection that can be optimized as needed to accommodate time constraints or reduce hands-on time. With this workflow, both the target size of custom panel designs and the level of methylation have little to no impact on final Picard metrics. Thus, this system is suitable for a wide variety of researchers interested in methylation detection.

REFERENCES

1. Vaisvila R, Pommaluri V K C, Sun Z, Langhorst B W, Saleh L, et al. (2019). EM-seq: Detection of DNA Methylation at Single Base Resolution from Picograms of DNA. BioRxiv. https://www.biorxiv.org/content/10.1101/2019.12.20.884692v1.full.pdf (accessed 10th Feb 2021).

2. Tanaka K & Okamoto A (2007). Degradation of DNA by bisulfite treatment. Bioorganic & Medicinal Chemistry Letters, 17(7), pages 1912–1915.

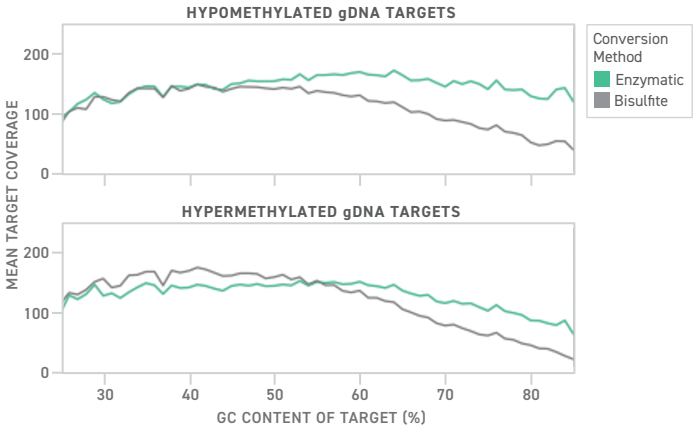


Figure 6. Coverage by Target GC Content for Hypo- and Hypermethylated gDNA Libraries Prepared with Enzymatic and Bisulfite Conversion Techniques. Enzymatic (teal) and bisulfite (grey) conversion library preparation methods were used to make libraries from hypo- and hypermethylated human control human cell lines. The Twist Targeted Methylation Sequencing protocol was performed using a custom 1.5 Mb panel and a single-plex reaction. Twist recommended capture conditions, including 2 µl of Methylation Enhancer, a Wash Buffer 1 temperature of 65°C, and a 2-hour hybridization time were used in each reaction. Sequencing was performed with a NextSeq 500/550 High Output v2 kit to generate 2x151 paired end reads. Data was down-sampled to 250x aligned coverage relative to the panel target size, mapped using the Bismark Aligner, and analyzed using Picard Metrics with a mapping quality threshold of 20. For both hypo- and hypermethylated gDNA types, target coverage is more evenly distributed across all GC bins when using Twist's enzymatic library preparation approach

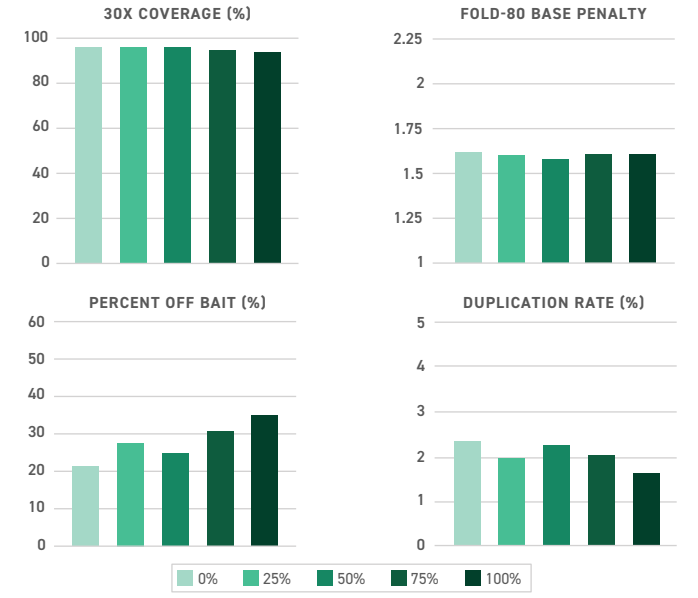


Figure 7. Picard HsMetrics for gDNA Libraries with Variable Methylation Levels. Libraries were prepared using the EM-seq conversion method and blends of hypo- and hypermethylated cell lines at ratios of 0, 25, 50, 75, and 100% methylation. A medium stringency designed 1 Mb panel was used to capture each gDNA library type in the Twist Targeted Methylation Sequencing protocol. Sequencing was performed with a NextSeq 500/550 High Output v2 kit to generate 2x151 paired end reads. Data was down-sampled to 250x aligned coverage relative to the panel target size, mapped using the Bismark Aligner, and analyzed using Picard Metrics with a mapping quality threshold of 20. Key hybrid selection metrics are steady for each gDNA library type, despite differences in CpG methylation levels.