

Enzymatic Methyl-seqおよびTwistターゲットエンリッチメントを使用した高感度メチル化検出

はじめに

DNA メチル化は、遺伝子発現を調節することにより、細胞増殖において重要な役割を果たします。場合によっては、異常なシトシンのメチル化が、遺伝子発現の発がん性の変化につながる可能性があります。歴史的には、患者検体でこれらの修飾を同定することは困難でした。亜硫酸水素ナトリウムで処理された DNA の次世代シーケンス (NGS) が、メチル化シトシン検出の効率的な手段として提供されていますが、化学的な変換プロセスがダメージを与え、GC に偏りのあるカバレッジおよび複雑性の低下につながる可能性があります¹。したがって、そのようなプロトコルは感度に限界があります。

Twist ターゲットメチル化のワークフローは、様々な状況でのメチローム解析のために高度に複雑で均一なライブラリ (リキッドバイオロジーによるがんスクリーニングのような新規のライブラリを含みます) を調製する完全なソリューションを導入しています。End-to-end のプロトコルは、Twist カスタムメチル化パネルを使用し、革新的な酵素による変換のプロセスと、適合性が高く迅速なハイブリッドキャプチャプロセスとを組み合わせることで実現します。このテクニカルノートでは、このワークフローの既存のプロトコルに勝る利点を明らかにし、ワークフロー全体を通してコントロールを使用することについて考察し、パネルターゲットのサイズおよびゲノム DNA メチル化レベルが下流のシーケンス指標に及ぼす影響を論証します。

結果

メチル化検出のためのライブラリ調製

従来のプロトコルでは、非メチル化シトシンをバイサルファイト処理によって化学的にウラシルに変換することにより、メチル化シトシン残基を同定します。PCR 後、非メチル化シトシンはチミンとしてシーケンスされ、メチル化シトシンはシトシンとしてシーケンスされます (図 1)。

しかしながら、バイサルファイト処理はしばしば DNA の切断を引き起こし、その結果、下流のサンプル調製、最終的にメチル化検出を複雑にする可能性があります²。Twist Bioscience® は、New England Biolabs® との提携により、Twist Targeted Methylation Sequencing Workflow の一部として NEBNext® Enzymatic Methyl-seq (EM-seq™) を提供しています。この革新的なプロセスが、化学的な変換の苛酷さなしにバイサルファイト処理と同じ変換の結果を達成し、優れた最終結果をもたらします。

メチル化検出のためのライブラリ調製は、6 つの重要なステップから成ります。

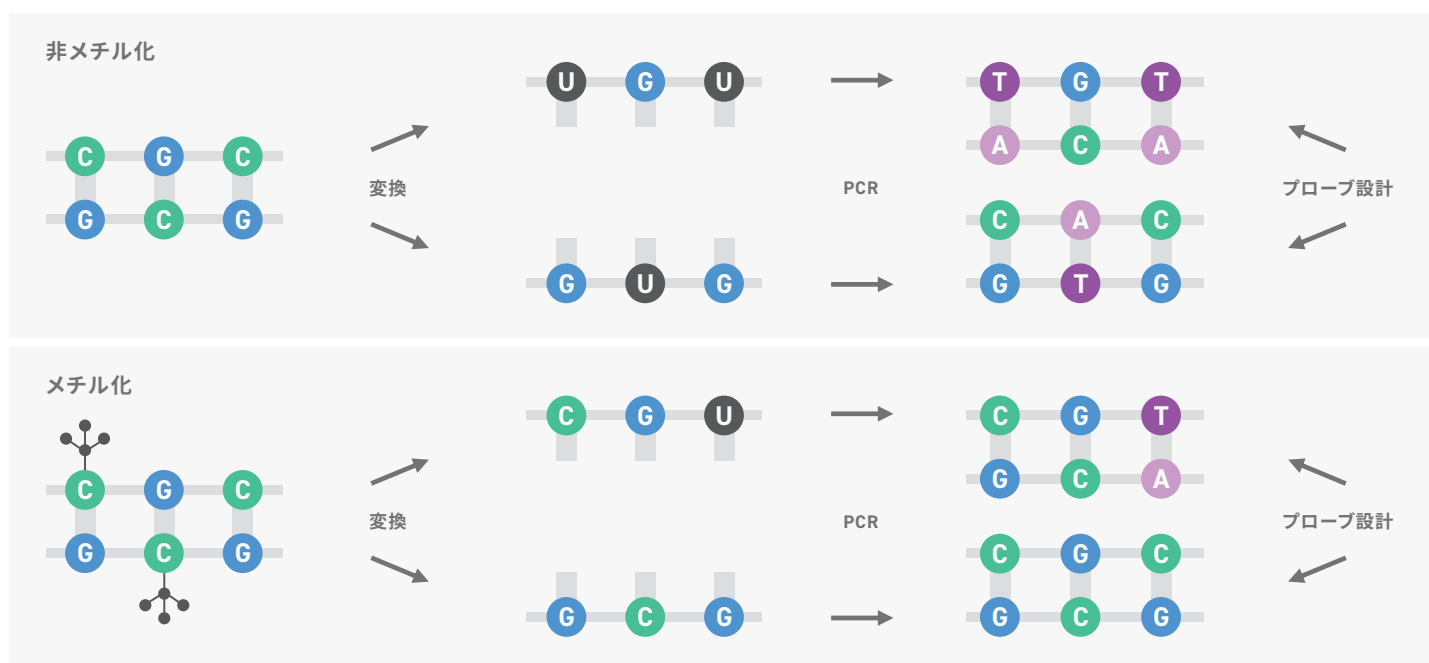
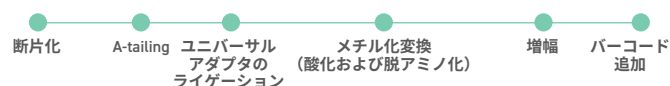


図 1: メチル化変換。メチル化シーケンスは、メチル化シトシンは完全なままで、非メチル化シトシンを脱アミノ化を介してウラシルに変換するための酵素的または化学的方法を含みます。増幅中、ウラシルは相補鎖上のアデニンと対になり、その結果、非メチル化シトシンの元の位置にチミンが組み込まれます。最終生成物は非対称であり、変換後、2 つの異なる二本鎖 DNA 分子が得られます (上段)。メチル化 DNA に対する同じプロセスでは、さらに追加のシーケンスセットが得られます (下段)。

酵素法とバイサルファイト法は、同等の効率で非メチル化シトシンをチミンに変換します (図 2)。ただし、酵素的に変換されたライブラリは、高 GC 領域で、より広く、より均一なカバレッジを示しています (図 3)。これは、これらの部位での DNA 切断が少ないことの結果であろうと推測されます。

Twist メチル化システムは、ライブラリ調製および酵素変換の前に物理的な断片化を利用することで、断片化フラグメントサイズに柔軟性をもたせています。Twist Bioscience では、メチル化検出のためにこのライブラリ調製法を使用する場合、ゲノムマテリアルを 200 ~ 300 bp 前後に断片化して、最終的に 350 ~ 450 bp の断片化フラグメントサイズとなるように調製するよう推奨しています。ただし、実際の断片化フラグメントサイズは、インプットのマテリアルの当初の完全性に依存します。ライブラリの収率は、インプットマテリアルの量と、ライブラリ調製時に使用される PCR 増幅サイクルの数に依存します。Twist Bioscience では、200 ng のインプットマテリアルと 9 サイクルの増幅から始めることを推奨しています。これらの推奨事項は、以後のハイブリッドキャプチャのために十分な材料が確実に使用できるようにするためのものです。最後の品質管理ステップで、フラグメント長の分布を確認します (図 4)。

Twist Targeted Methylation Sequencing Workflow におけるコントロール DNA の使用

不完全な変換はアッセイの偽陽性率を増大させます。これは変換されていない非メチル化シトシンがメチル化されていると誤って解釈されるためです。これを軽減するために、既知のメチル化レベルの DNA コントロールを使用して変換率を確認することができます。Twist Targeted Methylation Sequencing Workflow の一部として、NEB EM-seq Methylation Library Preparation キットは、CpG メチル化 pUC19 DNA および非メチル化 Lambda DNA をオプションのコントロールとして含んでいます。いずれのコントロールも、既知のレベルのメチル化を有し、シーケンス後の変換率の正確な決定を実現します。これらのコントロールは、ターゲットエンリッチメントパネルでは相補的プローブを欠く可能性があるため、これらのコントロールに対してはハイブリッドキャプチャを実施するべきではありません。その代わりに、それらをハイブリッドキャプチャの完了まで保管し、その後、シーケンスのためにサンプルと一緒にプールする必要があります。

Twist ワークフローにおけるこれらのコントロールの使用を実証するために、NEBNext EM-seq メチル化ライブラリ調製プロトコルの付録 A に記載されているプロトコルを使用して、ライブラリを作製しました。48 マイクロリットルの各 DNA コントロールを 1 回の反応で混合し、高速真空濃縮器を使用して、混合物を乾燥させました。得られた乾燥 DNA を、pH 8.0 の 0.1X TE 50 µl 中で再懸濁し、ライブラリプロセスを実施しました。

表 1 に、変換効率およびシーケンス後のメチル化レベルの実測値と期待値を示します。EM-seq は、両方のコントロールについて、99.5% を超える変換率で期待される効率を満たしました。非メチル化 Lambda DNA コントロールおよび CpG メチル化 pUC19 DNA コントロールの CpG メチル化レベルの期待値は、それぞれ 0.5% と 95 ~ 98% です。両方のコントロールについて、CpG メチル化レベルの測定値は期待されたレベルと一致しました。メチル化のレベルはメチル化のコールを行うツールを用いて計算されました。これらのデータは、既知のメチル化レベルの DNA コントロールを使用することで、変換プロセスの完全性とアッセイの偽陽性の最小化の確保を示唆するものです。

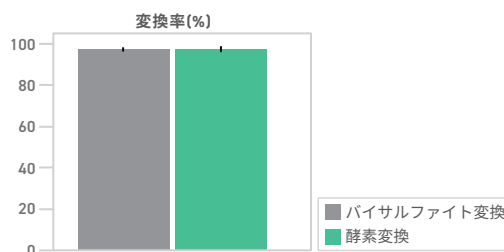


図 2: バィサルファイト変換と酵素変換の変換率比較。両方のライブラリ変換法とも、99.5% を超える変換率 (非 CpG 部位でチミンに変換されたシトシンの割合 (%) により測定) を実現しています。

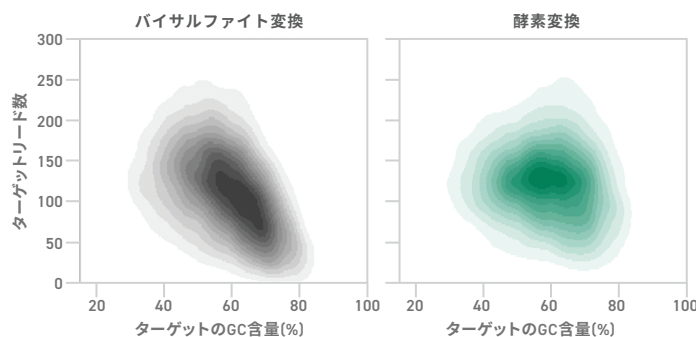


図 3: バィサルファイト変換および酵素変換のターゲット GC 含量別カバレッジ。いずれのライブラリ変換手法も Twist 高速ハイブリダイゼーションターゲットエンリッチメントシステムと適合性がありますが、酵素変換法で調製されたライブラリで、ハイブリッドセクションメトリクスの改善が認められます。バイサルファイト変換法 (グレー) を使用した場合、高 GC ターゲット領域がカバレッジの低下に関連します。一方、酵素変換法 (グリーン) を使用した場合、バイアスの程度の軽減が認められます。

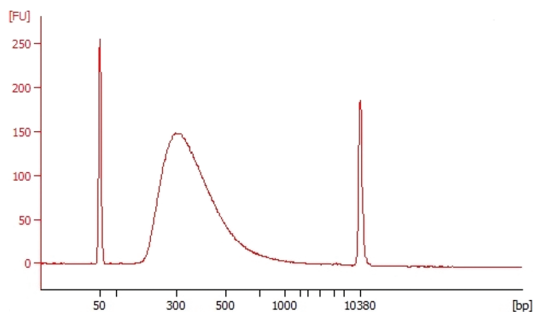


図 4: 最終のフラグメントの分析トレース。ライブラリ調製のために EM-seq 変換法を使用した最後の品質管理ステップ。平均ピーク長は約 375 bp です。

指標	非メチル化 LAMBDA DNA	CpGメチル化 pUC19 DNA
変換効率の期待値	>=99.5%	>=99.5%
変換効率の実測値	99.77%	99.57%
CpGメチル化レベルの期待値	最大0.5%	95~98%
CpGメチル化レベルの実測値	0.22228%	95.7572%

表 1: 変換効率および CpG メチル化レベルの期待値 vs. 実測値。CpG メチル化 pUC19 DNA コントロールおよび非メチル化 Lambda DNA コントロールを EM-seq で変換した場合の変換効率および CpG メチル化レベルの結果。

Twist では、EM-seq ワークフローにおける酵素変換の効率について懸念がある場合にこれらのコントロールを使用することを推奨しています。反応数 96 のキットに、各々のコントロールが十分にあり、別々にコントロールライブラリを 2 回調製できます。

メチル化検出用 Twist ターゲットエンリッチメント

Twist Targeted Methylation Sequencing Workflow ターゲットメチル化シーケンシングワークフローは、さらに Twist Fast Hybridization Target Enrichment System を使用して、高感度のカスタマイズされたメチル化検出を実現します。このシステムは、これまでの Twist ワークフローより短いハイブリダイゼーション時間で、シーケンス指標に悪影響を及ぼすことなくハイブリダイゼーションが平衡に達します。

Twist Fast Hybridization Target Enrichment System は、5 つの主要なステップを含みます。



このシステムにより、スループットの向上およびハンズオン時間が短くピペット使用も少ないマルチプレックス（最大 8 プレックス）が可能になります。Twist Targeted Methylation Sequencing Protocol は、様々なカスタムメチル化パネルと適合性のある最も良い結果につながる方法を推奨しています。ただし、あらゆるターゲットキャプチャシステムと同様、ターゲット領域のサイズやメチル化レベルなど、複数の要因が最終的なキャプチャデータの品質に影響を及ぼす可能性があります。以下では、これらの要因が下流のシーケンス指標に及ぼす影響について考察します。

ターゲット領域のサイズ

カスタムターゲット領域に関連する多くの要因が最終的なターゲットシーケンス指標に影響を及ぼします。最良の性能を実現するために、最適化が必要になることがあります。例えば、非常に小さいパネルデザイン (<0.5 Mb) やターゲット領域の GC 含量が高いパネルは、パネルデザインの軽微な変更特に敏感です。ターゲットをどれだけ含めるかは、オフターゲットとのトレードオフであり、最適なコントロールは、ターゲット領域の特性とパネルの意図された用途に依存します。例えばパネル設計プロセスで、中程度のサイズのパネルと少数のサンプルを扱う場合は、オフターゲットキャプチャの増加分を

相殺するために追加のシーケンスが必要であっても、特定のプローブの保持を選択する可能性があります。対照的に、はるかに小さいパネル（オフターゲットキャプチャにより、必要なシーケンス量が他のパネルより急激に増加）または非常に多数のサンプル（コストの増加が急速に累積される可能性）を扱う場合は、より厳格な設計条件を適用してコストの最適化を選択する可能性があります。

パネルサイズとシーケンス指標との関係性を評価するために、3 つの異なるパネルが Twist Targeted Methylation Sequencing Workflow で使用されました。広範なメチル化ターゲットに対し、0.5 Mb、3 Mb および 50 Mb のパネルサイズを網羅しました。今回の検討で使用した最も大きいパネルでは、7% に近いオフターゲットレベルとなり、すべてのパネルで 10% 未満のオフターゲットレベルとなりました。キャプチャの均一性 (Fold-80 ベースペナルティ) は、すべてのターゲットサイズについて並外れた値、すなわち 1.4 ~ 1.7 に達しました。30X カバレッジのプローブの割合は、すべてのパネルにおいて 90% を超えました。ただし、パネルサイズの増加に伴う duplicate 率の増加は、鍵となるどの指標を優先すべきかを定める際の検討事項となります。

多様なターゲット領域メチル化レベルの同定

メチル化レベルはゲノム内で差があります。特定のがんの早期発見のために異なるメチル化レベルを使用できることから、メチル化の検出のために使用するプロトコルがカスタムパネルのデザインと高度に適合し、かつ、高メチル化領域および低メチル化領域を同定する能力を持つことが極めて重要です。特定の塩基の変換は配列複雑性の低下につながり、下流のハイブリッドキャプチャステップで課題となります。ただし、これらの問題は、ライブラリ調製試薬、ハイブリッドキャプチャ試薬およびカスタムパネルデザインを適切に組み合わせることによって軽減することができ、その結果、AT/GC 含量およびメチル化レベルが異なる領域にまたがって、均一に分布するプローブカバレッジが得られます。ターゲット methyl-seq 用の Twist カスタムパネルは、1 ターゲット当たり 4 種のプローブで構成され、変換後、センスおよびアンチセンス DNA の両方のメチル化型および非メチル化型を最適にキャプチャするようになっています (図 1、右側を参照)。

Twist Targeted Methylation Sequencing Workflow を用いて得られるカバレッジの均一性を確認するために、1.5 Mb カスタムパネルと、2 つの異なる変換システム (EM-seq およびバイサルファイト処理) を

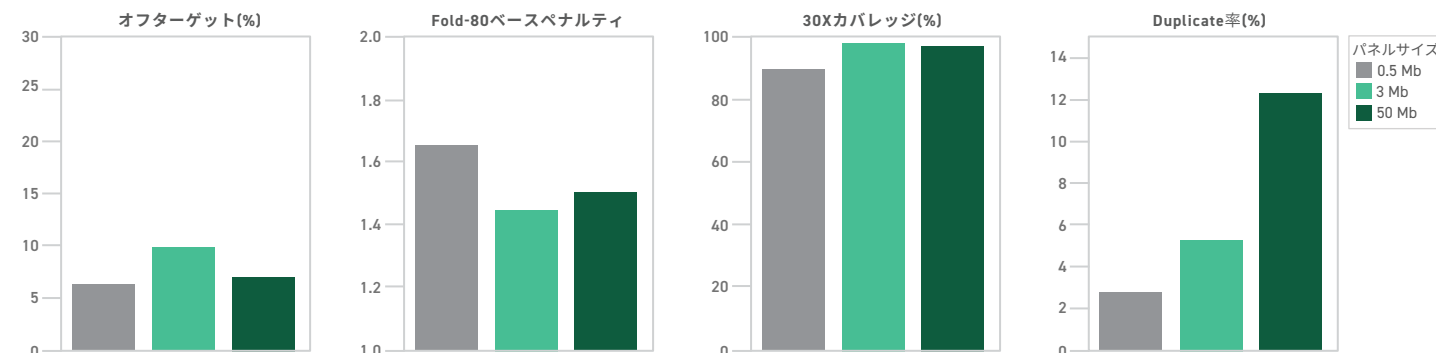


図 5: カスタムパネル内のターゲット領域のサイズおよび Picard 指標との関係。Twist Targeted Methylation Sequencing のプロトコルおよびシングルプレックス反応を使用する、0.5 Mb、3 Mb および 50 Mb のターゲットサイズを対象とするメチル化パネルの重要なキャプチャ指標。Twist が推奨するキャプチャ条件 (2 uL のメチル化エンハンサー、65°C の洗浄バッファー 1 の温度、2 時間のハイブリダイゼーション時間など) を各反応で用いました。NextSeq® 500/550 High Output v2 キットを用いてシーケンスを実行し、結果として 2x76 のペアードエンドリードを生成しました。データは、パネルのターゲットサイズに対して 200x のアライメントされたカバレッジになるようにダウンサンプリングし、Bismark Aligner を用いてマッピングを行い、Picard 指標とマッピングクオリティ閾値 20 を用いて解析しました。

用いて、高メチル化および低メチル化ゲノム DNA インプットマテリアルを併用して、プローブカバレッジプロットを作成しました。図 6 は、EM-seq 変換法（ティールブルー）を使用した場合の両方のメチル化レベルについて、ターゲットリード数の均一な分布を示しています。対照的に、業界をリードするバイサルファイト変換プロセス（グレー）のターゲットリード数は、前者と比較して不均一な結果となりました。Twist カスタム パネル デザイン は、Twist Targeted Methylation Sequencing Workflow と併用することで、ターゲットメチル化レベルに関係なく、あらゆるターゲット DNA 種を効率的にキャプチャします。

Twist Twist Targeted Methylation Sequencing Workflow の性能を様々なメチル化レベルにまたがってより詳しく調べるために、1 Mb パネルを使用して様々なメチル化レベルのライブラリを調製し、キャプチャしました。これらのライブラリは、低メチル化および高メチル化コントロールヒト細胞株 (EpiScope® Unmethylated HCT116 DKO gDNA、品番 3521 および EpiScope Methylated HCT116 gDNA、品番 3522; Takara Bio USA) を混合して、メチル化率 0%、25%、50%、75% および 100% のライブラリを調製しました。図 7 は、重要なキャプチャ指標を示しており、メチル化レベルと最終的な Picard 指標との関係がないことを明らかにしています。これらのデータは、Twist メチル化システムが低メチル化および高メチル化 DNA 領域のいずれにも適合性があることを実証しています。

結論

Twist Bioscience は現在、NEB から提供される酵素変換とカスタムメチル化パネルおよび最適化されたターゲットエンリッチメント試薬を組み合わせた end-to-end のターゲットメチル化検出ワークフローを提供しています。結果として、Twist Targeted Methylation Sequencing Workflow は、時間的な制約に対応したり、ハンズオン時間を削減したりといった必要に応じての最適化が可能な高感度のメチル化検出を実現します。このワークフローを使用することで、カスタムパネルデザインのターゲットサイズもメチル化のレベルも、最終的な Picard 指標にほとんど影響を及ぼしません。したがって、このシステムは、メチル化検出に関心を持つ様々な研究者に適しています。

参考文献

1. Vaisvila R, Pommali 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.

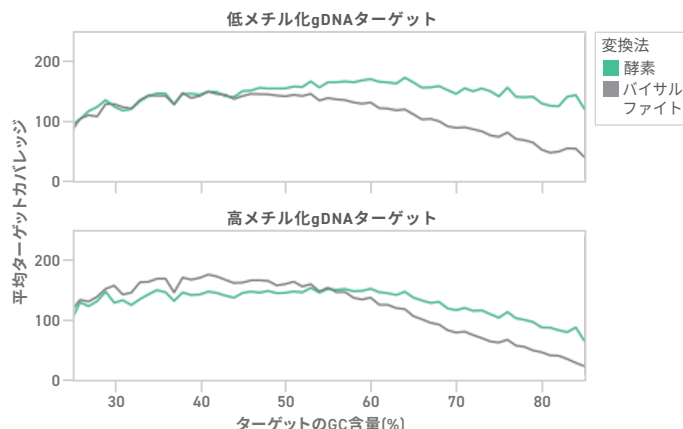


図 6：酵素変換法およびバイサルファイト変換法を用いて調製された低メチル化および高メチル化 gDNA のターゲット GC 含量別カバレッジ。酵素変換ライブラリ調製法（ティールブルー）およびバイサルファイト変換ライブラリ調製法（グレー）を使用して、低メチル化および高メチル化コントロールヒト細胞株からライブラリを調製しました。カスタム 1.5 Mb パネルとシングルプレックス反応を使用して、Twist Targeted Methylation Sequencing のプロトコルを実施しました。Twist が推奨するキャプチャ条件（2 μ L のメチル化エンハンサー、65°C の洗浄バッファー 1 の温度、2 時間のハイブリダイゼーション時間など）を各反応で用いました。NextSeq 500/550 High Output v2 キットを用いてシーケンスを実行し、結果として 2x151 のペアードエンドリードを生成しました。データは、パネルのターゲットサイズに対して 250x のアライメントされたカバレッジとなるようダウンサンプリングし、Bismark Aligner を用いてマッピングを行い、Picard 指標とマッピングクオリティ閾値 20 を用いて解析しました。Twist の酵素変換ライブラリ調製法を使用した場合、低メチル化および高メチル化 gDNA 型の両方について、ターゲットカバレッジはすべての GC 区間にまたがってより均一に分布しています。

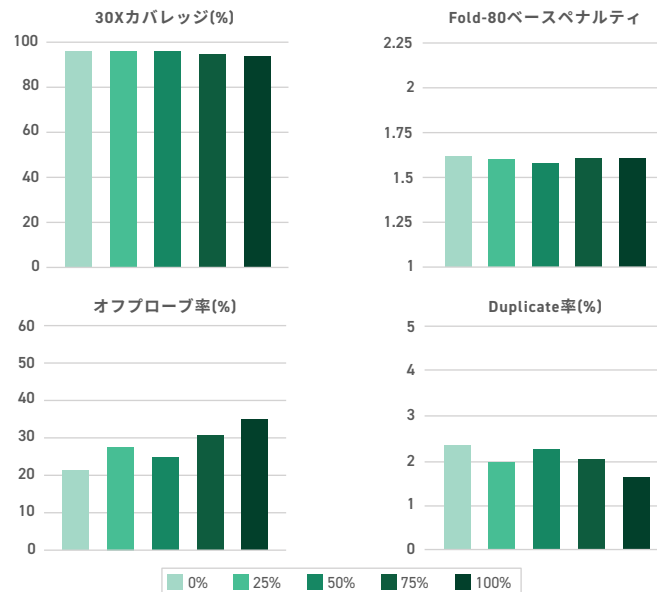


図 7：メチル化レベルが異なる gDNA ライブラリ用の Picard HsMetrics。EM-seq 変換法と、メチル化率 0%、25%、50%、75% および 100% の低メチル化および高メチル化細胞株の混合物を用いて、ライブラリを調製しました。Twist Targeted Methylation Sequencing のプロトコルを用いて各レベルの gDNA ライブラリをキャプチャするために、中程度の厳密性で設計された 1Mb パネルを使用しました。NextSeq 500/550 High Output v2 キットを用いてシーケンスを実行し、結果として、2x151 のペアードエンドリードを生成しました。データは、パネルのターゲットサイズに対して 250x のアライメントされたカバレッジになるようダウンサンプリングし、Bismark Aligner を用いてマッピングを行い、Picard 指標とマッピングクオリティ閾値 20 を用いて解析しました。CpG メチル化レベルに差があるにもかかわらず、各レベルの gDNA ライブラリの重要なハイブリッドセレクション指標は一定です。