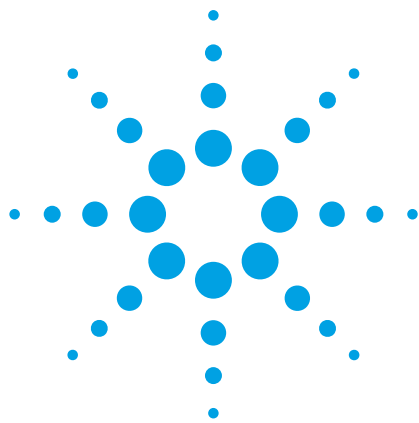




Agilent SureSelect^{XT} Human Methyl-Seq を用いた 1 塩基の解像度での DNA メチル化定量解析

技術概要

要旨

Agilent SureSelect^{XT} Human Methyl-Seq は、ヒトゲノムの低メチル化および高メチル化シトシン部位を 1 塩基の解像度で解析できる非常に強力なツールです。このアッセイは、最先端のターゲットエンリッチメントプラットフォーム SureSelect と、DNA メチル化研究のゴールドスタンダードで包括的な検出システムであるバイサルファイトシーケンシングを組み合わせたものです。このツールを利用することで、エピジェネティクス研究に最も関係する領域のみを、これまでにない深いカバレッジでシーケンスすることが可能となりました。がん、インプリンティング異常、行動障害、および精神疾患などの幅広い疾患に、これらの領域のメチル化状態が関連していると考えられています。

Agilent SureSelect^{XT} Human Methyl-Seq は、370 万以上の CpG ジヌクレオチド配列のメチル化状態について網羅的に解析することを可能としました。Human Methyl-Seq は、プロモーター、CpG アイランドと定義される領域、および近年 shore および shelf と呼ばれている CpG アイランドの両側 4000 塩基対までの領域をターゲットとしています。多くのメチル化の変化はプロモーターや CpG アイランドにはなく、そのほとんどが 2kb 以内の CpG アイランド shore にあることが研究結果から示されています。SureSelect^{XT} Human Methyl-Seq は、既知の DNA メチル化状態が異なる領域 (differentially methylated regions; DMRs) もターゲット領域としています。



Agilent Technologies

はじめに

遺伝子発現を制御するエピジェネティクスの関連性が、現在、ゲノミクス分野、特にがん研究において非常に注目されています。エピジェネティクス研究のためにいくつかの手法が開発されてきましたが、既存の手法には解像度およびバイアスがかかりやすい点で限界があります。

- メチル化解析マイクロアレイでは、メチル化領域の濃縮と、シーケンシングではなくハイブリダイゼーションを組み合わせています。この手法は解像度が低く、個々の CpG 部位のメチル化状態を解析することはできません。
- 全ゲノムバイサルファイトシーケンシング (Whole genome bisulfite sequencing; WGBS) では、全ゲノムのバイサルファイト処理後にシーケンシングを行います。この手法は大規模な研究では費用と時間がかかり過ぎるため、プロファイリングやバイオマーカー探索のために多数のサンプルを解析する研究にとっては現実的ではありません。
- MeDip-Seq では、メチル化領域を濃縮した後、シーケンシングを行います。この手法は、抗体を用いて、メチル化した 1 本鎖 DNA 断片を免疫沈降させた後にシーケンシングします。メチル化したターゲットの濃縮が可能ですが、繰り返し配列と CpG-rich な配列に対してバイアスがかかります。

- Reduced representation bisulfite sequencing (RRBS) では、ゲノムの一部分についてメチル化解析を行うことで、シーケンス量を減らす手法です。DNA はまず制限酵素により CpG 認識部位で切断されます。その後アダプタ付加、ゲルによるサイズ選択、バイサルファイト変換と PCR を行い、DNA を濃縮します。この濃縮法では特定のメチル化領域を選択することはできず、繰り返し配列と CpG-rich な配列にバイアスがかかります。

SureSelect^{XT} Human Methyl-Seq は 1 塩基の解像度で DNA メチル化の定量解析を可能にする新しい手法を提供します。SureSelect Target Enrichment System を用いることで、興味のあるゲノム領域のみをシーケンスすることが可能となり、

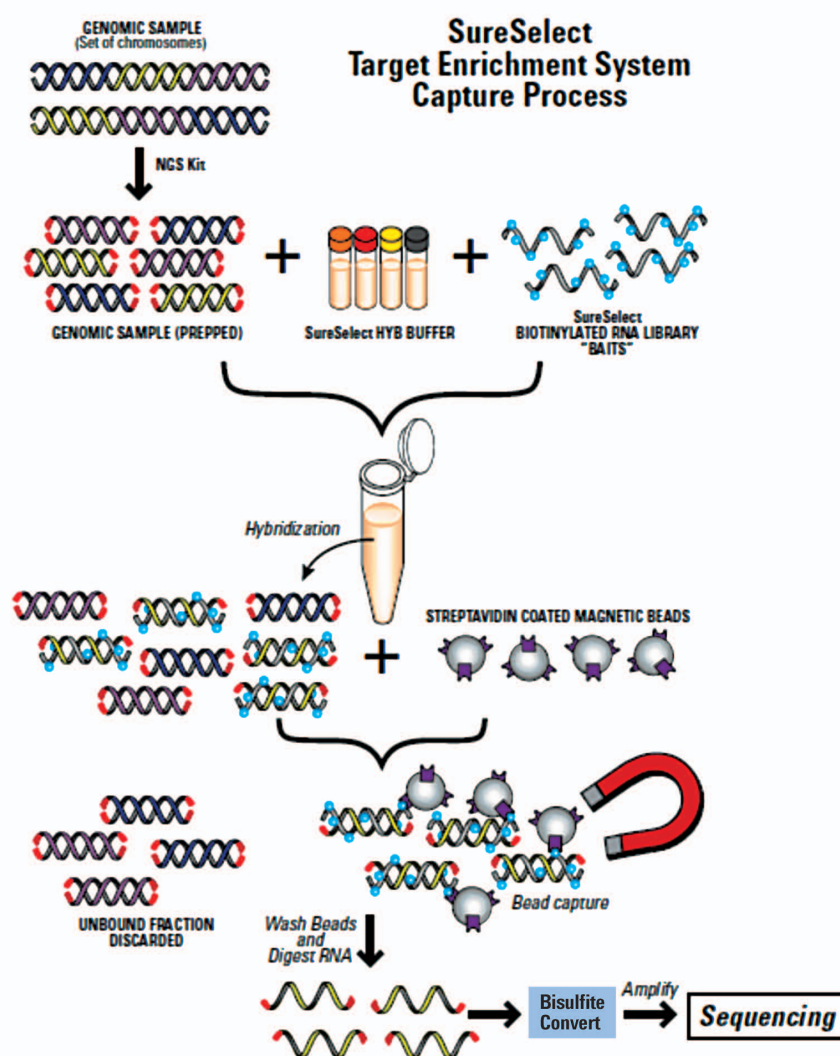


図 1. SureSelect ターゲットエンリッチメントシステムワークフロー

研究者は作業効率を高め、費用を減らし、解析できるサンプル数を増やすことができます。その結果、より容易に、統計学上妥当なサンプル数をそろえた実験を行う事ができます。SureSelect で用いるプローブはメチル化状態に依存しないため、あらゆるターゲット領域をメチル化状態によらずキャプチャすることができます。SureSelect^{XT} Human Methyl-Seq の包括的なターゲット領域には、CpG アイランド、Gencode プロモーターに加えて、CpG アイランドの shore と shelf、DNase I 高感受性部位、Refseq 遺伝子、Ensembl regulatory features にある既知の組織・腫瘍特異的 DMRs が含まれます。

デザインコンテンツ一覧 – 84 Mb デザイン、370 万もの CpG 部位

- CpG アイランド
- がん・組織特異的 DMR
- Gencode プロモーター
- 次に含まれる DMR または regulatory features
 - CpG アイランド、shore および shelf ± 4 kb
 - DNase I 高感受性部位
 - Refseq 遺伝子
 - Ensembl regulatory features

SureSelect^{XT} Human Methyl-Seq と全ゲノムデータとの高い相関

ヒト胎児肺由来線維芽細胞 (IMR 90) を使ったデータは、論文で報告されたデータと高い相関 ($R > 0.93$) を示しました。図 2 の各六角形は、平均メチル化レベル (>10 リードのフィルタを設定) を示しています。IMR90 細胞より得られた SureSelect^{XT} Human Methyl-Seq データを、全ゲノムシーケンシング (whole genome sequencing; WGS) のデータ² の平均メチル化レベルと比較しています。

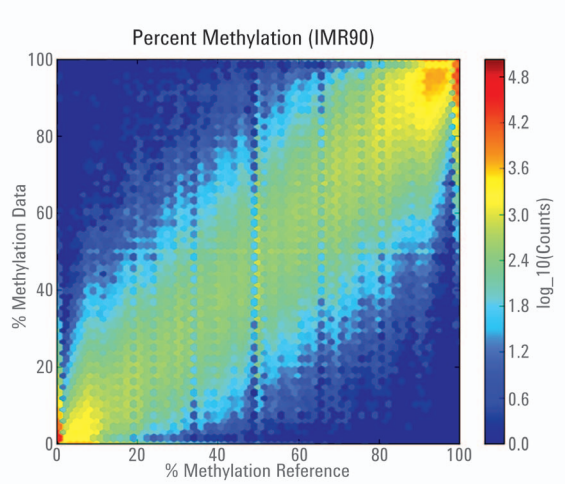


図 2. 胎児肺由来線維芽細胞の SureSelect^{XT} Human Methyl-Seq データと報告されている WGS のデータとの強い相関 ($R > 0.93$)

大腸がん細胞を用いたコンセプトの検証

SureSelect によるターゲットエンリッチメントを利用した高感度かつ高精度なメチル化検出による、HCT116 と、メチル基転移酵素 (DNMT1^{-/-} および DNMT3b^{-/-}) ダブルノックアウト細胞株との間の DNA メチル化の違いを示します。図 3 に示す通り、HCT116 ダブルノックアウト細胞株では DNA メチル化が劇的に減少しています。メチル基転移酵素はゲノム DNA のメチル化を行うため、予測される結果と一致しています。

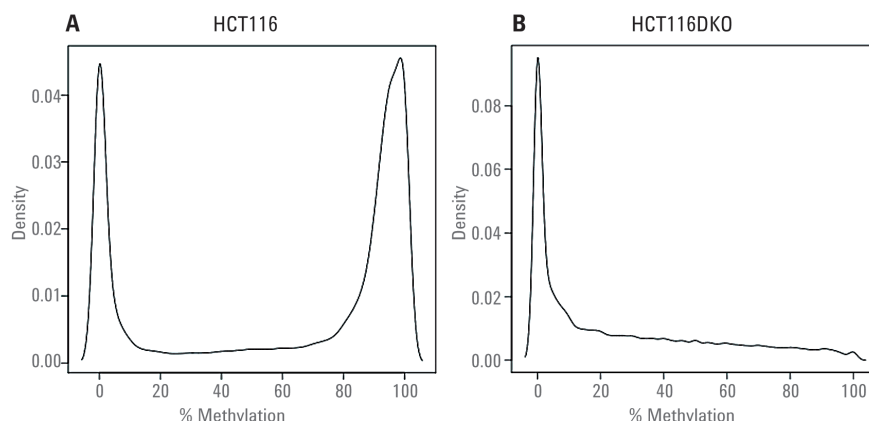


図 3. 野生型 HCT116 (A) とそのメチル基転移酵素ダブルノックアウト細胞株 (B) の間のメチル化の違い

高いカバレッジで新規メチル化領域を探索

Agilent のプローブはメチル化状態に依存しないので、ゲノム内の既知メチル化領域を全てターゲットとし、ターゲット領域のカバレッジを増加させることができます。カバレッジの増大により、シーケンス深度が低いと検出されないような少数のサブpopulationがもつ新規メチル化部位の検出が可能となります。SureSelect^{XT} Human Methyl-Seq の 4Gbp のユニークにマッピングされたリードは、全ゲノムバイサルファイトシーケンシングの 91Gbp のユニークにマッピングされたリードより、ターゲット領域においても CpG アイランド特異的に存在する CpG においても高いカバレッジを示しています。(図 4)

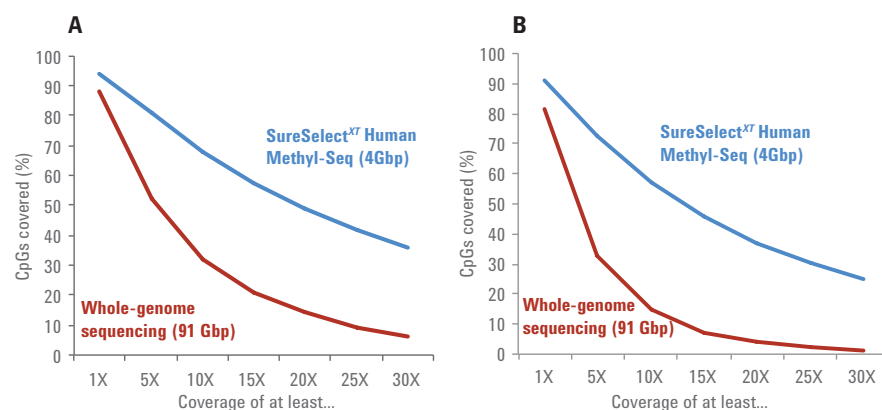


図 4. 全ゲノムシーケンシング (180 Gb の生シーケンスデータで 91 Gb のユニークにマッピングされたデータ) と SureSelect^{XT} Human Methyl-Seq (10 Gb の生シーケンスデータで 4 Gb のユニークにマッピングされたデータ) において、ターゲット領域にある約 370 万の CpG (A) と、CpG アイランドにある約 190 万の CpG (B) の %CpG カバレッジ。

様々な組織に対して差のあるメチル化状態を探索

SureSelect^{XT} Human Methyl-Seq を用いて、5 種類の正常組織における組織特異的 DMRs を検出しました。平均メチル化レベルは 200 bp ウィンドウで算出し、クラスター解析しました。その結果、208,000 の共通メチル化領域と 111,000 の非メチル化領域が検出されました。図 5 に各組織で検出された組織特異的なメチル化の状態を示します。

1 塩基の解像度でのがん特異的 DMR

腫瘍特異的 DMRs を正常組織と大腸がん組織で比較しました。SureSelect^{XT} Human Methyl-Seq を用い、既知の DMRs だけでなく、新規 DMRs の可能性がある領域を 1塩基の解像度で検出しました。comprehensive high-throughput arrays for relative methylation (CHARM) とパイロシーケンシングによって特定された HOXA3 遺伝子にある DMRs の存在 (正常組織で 52 %, 腫瘍組織で 72 %)を Methyl-Seq で確認した例を示します。加えて SureSelect^{XT} Human Methyl-Seq を用いることにより、HOXA3 プロモーター、HOXA3/HOXA4 の遺伝子内・遺伝子間領域にさまざまな DMR が検出されました。(図 6)

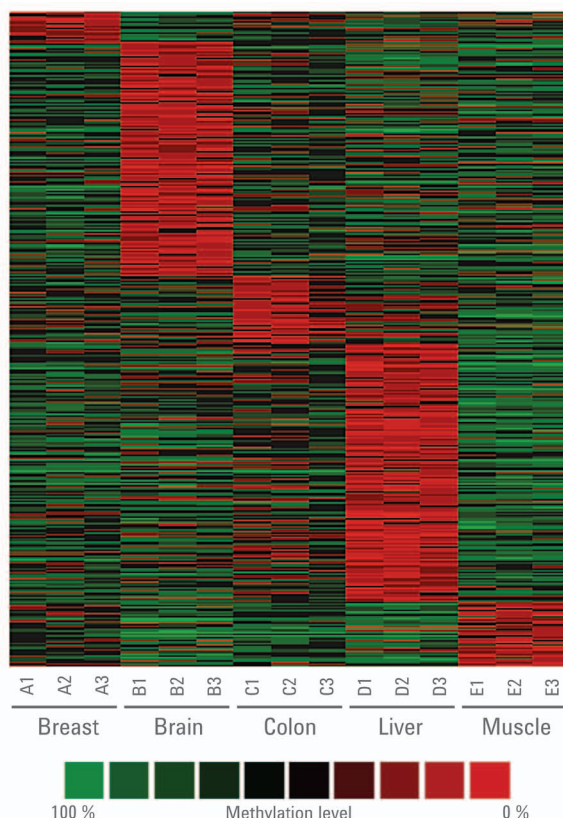


図 5. SureSelect^{XT} Human Methyl-Seq を用いて 5 種類の組織で特定された組織特異的なメチル化領域



図 6. 正常組織 (上) と大腸がん (下) における腫瘍特異的 DMR の比較。既知および新規 DMR が 1 塩基の解像度で示されています。

データ解析

SureSelect^{XT} Human Methyl-Seq の目的は、DNA バイサルファイト処理と、ターゲット領域に対するハイスループットシーケンシング結果を組み合わせ、1塩基の解像度で個々のシトシンのメチル化状態を検出することにより、細胞のエピゲノム状態全体を把握することです。溶液中でのハイブリダイゼーションによるキャプチャの後、データ解析を行う前に前処理段階があります。この前処理段階には、DNA のバイサルファイト処理、バイサルファイト処理した (bisulfite treated: BST) DNA のハイスループットシーケンシング、複数のサンプルを解析しているときには demultiplex 作業が含まれます。

Agilent では現在、SureSelect^{XT} Methyl-Seq のデータ解析用のツールを提供、サポートしていませんが、参考情報として、データ解析パイプラインの一例を下記に示します。このデータ解析パイプラインは、アラインメント、duplicate 除去、正規化、検出、メチル化 % 算出、およびメチル化部位の特定の一連の段階から構成されます (図 7)。

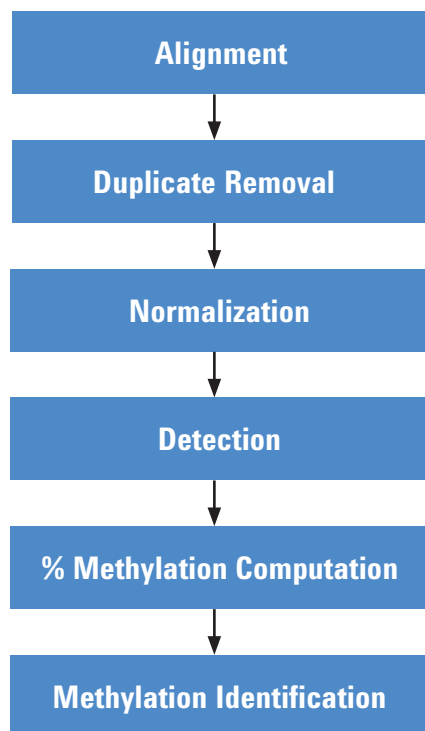


図 7. SureSelect^{XT} Methyl-Seq データ解析ワークフロー

1. アラインメント

DNA をバイサルファイト処理 (BST) すると、メチル化したシトシンはそのまま、メチル化していないシトシンはウラシルに変換されます。この BST DNA の特徴によりメチル化シトシンを検出することができますが、それと同時に複雑性が増し、得られたシーケンシングリードのリファレンスゲノムにおける正しい位置を特定することがより困難になります。一般に、BST DNA リードのメチル化状態は、修飾をうけていないリファレンスゲノムと BST リードを比較することで推測されます。この比較は BST リードをゲノム配列にアラインメントすることで可能になります。

ここでは、アラインメントを行うために、バイサルファイトシーケンスのアラインメントとメチル化検出用のツールである Bismark を用いています (図 8)。アラインメント過程で、リード配列の C を T に変換したバージョンと、G を A に変換 (相補鎖での C-T 変換に相当) したバージョンが生成されます。リファレンスゲノムについても同様に変換したバージョンが生成され、リードがそれぞれ変換し

た配列にアラインメントされます。そうすることで、Bismark は最もミスマッチが少なかったものを、最も可能性の高いバイアスのないリードのアラインメントとして予測することができます。リードが複数のバージョンのゲノムにアラインメントされ、ミスマッチの数が等しい場合には、そのリードは破棄されます。その次に各リードはあらかじめ C を T に、G を A に変換したバージョンのリファレンスゲノムにアラインメントされます。ここで Bismark は、バイサルファイト処理によって導入されたバイアスにかかわらず、そのリードのゲノムでの正しい位置と strand を見つけようとします。この段階で、配列が複数の位置にアラインメントされ、最小となるミスマッチ数が同数の場合、その配列は破棄されます。同数ではない場合は、ミスマッチが最小のものが正しいアラインメントとして選ばれます。

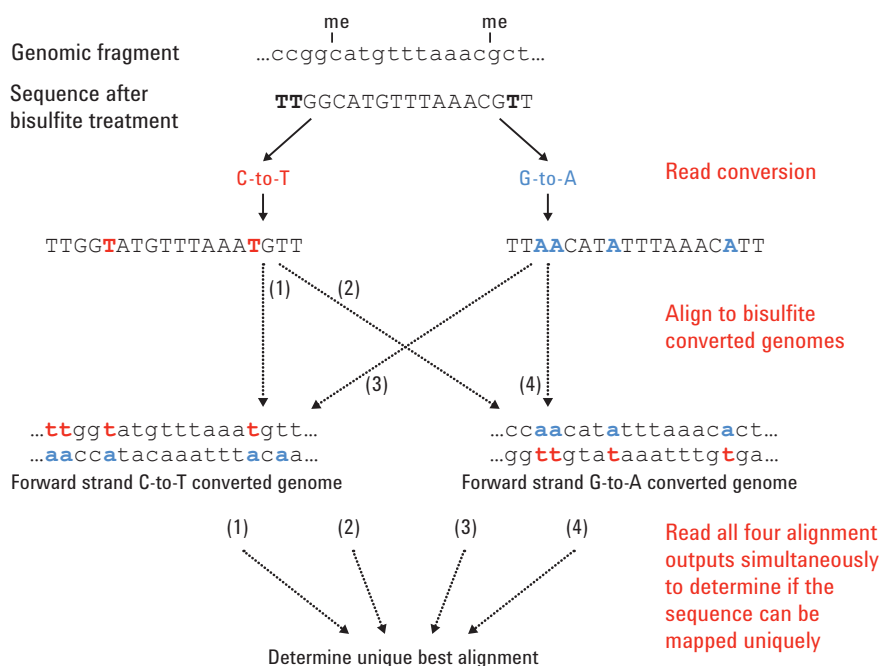


図 8. Bismark¹ バイサルファイトシーケンスアラインメントとメチル化検出のアプリケーション

2. duplicate 除去

PCR duplicate を除くために、Strand、染色体、およびゲノム上の位置がまったく同一のリードペアが複数ある場合には、1 つだけが残され、その 1 対のリードペアのみがその後の解析に用いられます。
(図 9)

3. 正規化

複数のサンプルをデータ処理する場合、ランダムにリードをサンプリングし、同じ数を割り当て、それらがお互いに比較可能な状態にします。これは解析した領域に渡って同程度の平均リード深度を持つようにするために実行されます。

4. 検出

アラインメントでリードのゲノム上の位置が決まったら、BST リード配列とリファレンスゲノム配列とを単純比較することにより、シトシンのメチル化状態が Bismark で決定されます (図 10)。このステップでは、オリジナルのリファレンスゲノム配列と BST リード配列を比較して、リードがマップされている側の Strand によって、C-T 変換 (または反対側の鎖では G-A 変換) を起こしていないシトシンをメチル化シトシンとしてマークします。

5. メチル化 % の算出

メチル化検出後、Bismark は、全シーケンシングリードに存在する各シトシンのメチル化状態をリストにまとめて、サマリレポートを出力します。ある特定のシトシンについてメチル化 % (% m) を計算するために、Bismark の Primary Output はゲノム上の位置に基づいて並べられており、メチル化 % は、特定の位置において検出されたシトシンの総数に対するメチル化シトシンの割合で計算されます (図 11)。

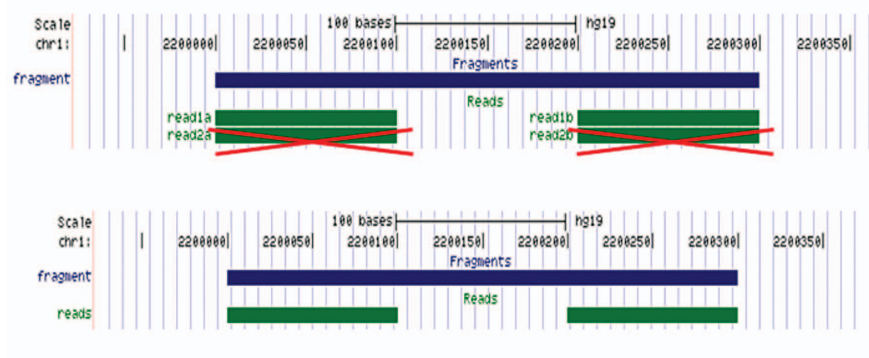


図 9. Duplication : PCR duplicate の除去

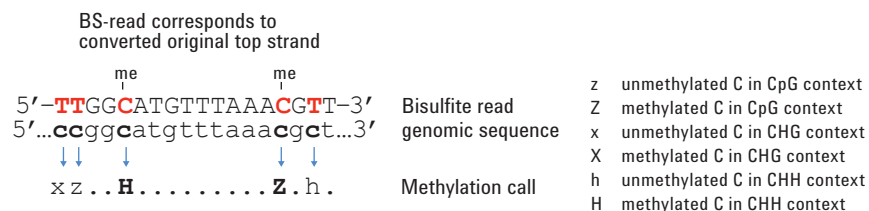


図 10. メチル化検出

6. メチル化の同定

特定のシトシンのメチル化状態を、さらにメチル化レベルの仮想分布と統計学的有意性に基づいて算出することができます²。

Bismark のオプションとヘルプに関する詳細な情報や資料については、ご使用の Bismark のバージョンに対応したオンラインユーザーガイドとリソースページを参照してください。(参考文献 1)

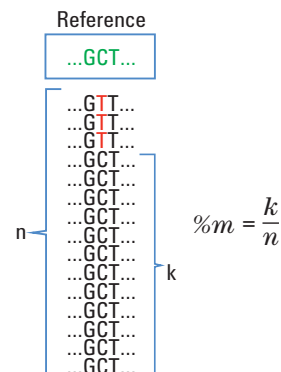


図 11. メチル化レベルの算出

結論

Agilent のSureSelect^{XT} Human Methyl-Seq ターゲットエンリッチメントプラットフォームは、ヒトメチロームにおいて関心がもたれている領域のみを濃縮してシーケンスする、包括的、効率的で、堅牢かつ費用対効果の高い方法です。SureSelect^{XT} Human Methyl-Seq ターゲットエンリッチメントを使用した後、バイサルファイト処理した DNA をシーケンスすることで、ターゲット領域の DNA メチル化状態を 1 塩基の解像度で解析することができます。ここで示した結果は、論文で報告されている全ゲノムデータと非常に相関が高く、その信頼性が高いことを示しています。さらにSureSelect^{XT} Human Methyl-Seq はマルチプレックスシーケンシングにおいて、再現性の高いエンリッチメント、リードの分布、およびシーケンスカバレッジを可能にします。また、SureSelect^{XT} Human Methyl-Seq は既知の組織および腫瘍特異的 DMRs を確認することばかりでなく、新規 DMR 候補を特定することも可能であり、ゲノムのメチル化領域を調べるために使われる既存の方法と比べても、さらなる研究への貢献が期待されます。

参考文献

1. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq Applications. *Bioinformatics*, **2011**, 27:1571-1572.
2. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic difference. *Nature*, **2009**, 462:315-322.

付録：メチル化データ解析の実践ガイド*

Genome Preparation : (このステップは各ゲノムに対して 1 度だけ行う必要があります)

USAGE: bismark_genome_preparation
[options] <path_to_genome_folder>

Sample command :

```
bismark_genome_preparation --path_  
to_bowtie /usr/local/bowtie/ --  
verbose/data/genomes/homo_sapiens  
/GRCh37/
```

Alignment :

USAGE: bismark [options]
<genome_folder> {-1 <mates1>
-2 <mates2> | <singles>}

典型的な 40 bp シーケンシングのシングルエンド解析のためのサンプルコマンド :

```
bismark -q --phred64-quals -n 1 -l 40  
--directional /data/genomes/homo_  
sapiens/GRCh37/ s_1_sequence.txt
```

duplicate 除去 :

The deduplicate_bismark_alignment_output.pl というヘルパースクリプトが利用できます。

Detection :

USAGE: ./methylation_extractor
[options] <fi lenames>

ペアエンドファイル用サンプルコマンド :

```
methylation_extractor -p --merge_non_  
CpG --comprehensive s_1_sequence.txt  
_bismark_pe.txt
```

% methylation computation :

The genome_methylation_bismark2bedGraph_v3.pl というヘルパースクリプトが利用できます。

<http://www.bioinformatics.bbsrc.ac.uk/projects/download.html#bismark>

* Bismark のオプションとヘルプに関する詳細な情報や資料については、ご使用の Bismark のバージョンに対応したオンラインユーザーガイドとリソースページを参照してください。

www.agilent.com/genomics/sureselect

本製品は研究用です。診断目的に使用する事はできません。
また、内容は予告無しに変更されることがあります。

アジレント・テクノロジー株式会社
© Agilent Technologies, Inc., 2012
Published in Japan, August 31, 2012
5991-0166JAJP



Agilent Technologies