

2種類のInfinium®ケミストリーを使用した HumanMethylation450 BeadChipによる 広範囲なカバレッジ

はじめに

HumanMethylation450 BeadChipは、業界で最も信頼性が高く、実績あるDNA解析プラットフォームであるイルミナのInfiniumアッセイを使用しており、エキスパートに選ばれた包括的なカバー領域と、サンプル数の多いエピゲノムワイド関連の研究に最適な高スループットを特長としたユニークな製品です。BeadChipは、Infinium I アッセイケミストリーとInfinium II アッセイケミストリーの技術を組み合わせることにより、RefSeq遺伝子の99%、CpGアイランドの96%、およびエキスパートコンソーシアムから求められたコンテンツのすべてをカバーしています。本テクニカルノートでは、この独自のテクノロジーの背景知識をご紹介しますとともに、BeadChipによって得られたデータ解析を補足するための情報を提供いたします。

2種類のInfiniumケミストリーによりカバー領域を拡大

Infiniumメチル化アッセイは、DNAサンプル内の個々のCpGサイトを調べるためにデザインされた、ターゲット特異的な長鎖プローブが結合したビーズを使用しています。DNAメチル化は、バイサルファイト処理したゲノムDNAの定量的ジェノタイプングにより測定されます。Infinium I アッセイおよびInfinium II アッセイが、お互いに補完的に働くことにより広いターゲット領域をカバーします。

Infinium I アッセイ

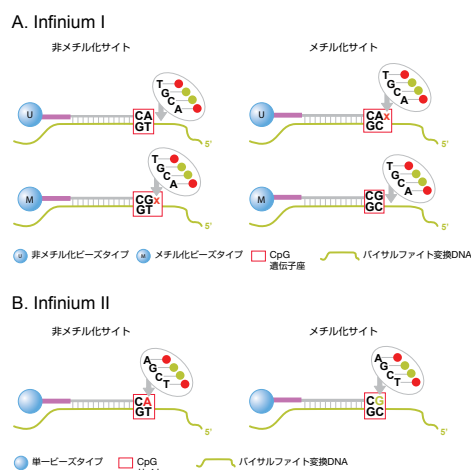
Infinium I アッセイでは、1つのCpGサイトに対して、1個の「非メチル化」DNA検出用プローブおよび1個の「メチル化」DNA検出用プローブの合わせて2個のプローブが利用されます（図1A）。各プローブの3'-末端は、保護されたシトシン（メチル化デザイン）、またはバイサルファイトによる変換と全ゲノムの増幅によって生成するチミン塩基（非メチル化デザイン）のいずれかにマッチするようにデザインされています。

Infinium I アッセイ用のプローブは、メチル化の状態は50bpの範囲内では局所的に相関するという仮説に基づき、メチル化潜在CpGサイトは、「メチル化」(C) または「非メチル化」(T) 検出サイトと一致したデザインとなっています。この共メチル化 (co-methylation) 理論は、染色体6、20、および22のバイサルファイトシーケンシングにおいて、50bpの範囲に存在するCpGサイトの90%以上が、同様なメチル化の度合を示していることを明らかにした研究より支持されています¹。また別の研究では、一般的に隣接サイトのメチル化の度合が相関する傾向にあることが示されており、この相関が細胞種または近接する多型配列に依存する可能性があることが示唆されています²。

Infinium II アッセイ

Infinium II アッセイは、1個のターゲットサイトに対して1個のプローブのみを使用するようにデザインされています（図1B）。プローブの3'-末端は、検出サイトのすぐ上流の塩基に

図1：Infinium I およびInfinium II アッセイデザインにより広範囲の領域をカバー



HumanMethylation450 BeadChipは、Infinium I およびInfinium II の両アッセイを使用し広範囲の領域をカバーします。

- Infinium I アッセイデザインは、1個のCpGサイトに対して、メチル化状態と非メチル化状態をそれぞれ1個ずつ、合計2個のビーズタイプを使用します。
- Infinium II デザインは、1種類のビーズタイプを使用し、メチル化サイトは、ハイブリダイゼーション後の1塩基伸長段階において決定されます。

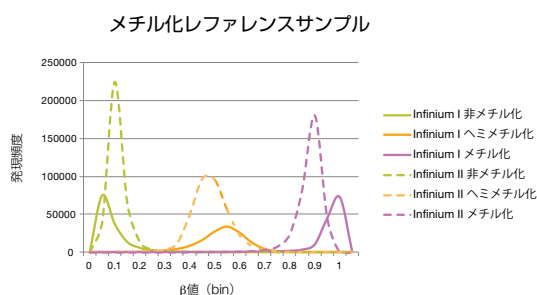
相補的で、1塩基伸長の結果、「メチル化」Cまたは「非メチル化」Tに相補的な、標識されたGまたはAの塩基がそれぞれ付加されます。

Infinium II アッセイは、1個の50merプローブを使用してメチル化状態を決定するため、「全か無か」の検出にはできません。しかし、縮重オリゴヌクレオチドプローブを用いることにより、50merの配列内に別のCpGサイトが存在したとしても、最大3箇所まで（合計27個の可能な組み合わせ）存在しても、データに影響を及ぼさないことが確認されています。この特長により、ターゲットに隣接するCpGサイトのメチル化状態とは関係なく、検出サイトでのメチル化状態を評価することが可能となります。さらに、各ターゲットサイトで必要なビーズは1種類のみであるため、検出可能なCpGサイト数が増加します。このため、Infinium II デザインは、多くの研究に幅広く利用されています。

Infinium I アッセイ vs. Infinium II アッセイ

Infinium I アッセイとInfinium II アッセイでは、そのケミストリーが異なるため、それぞれに特徴的な優位性を有しています。この2つのケミストリー間の相違は、データセット内の、それぞれに特徴的なβ値分布として観察されます。図2には、β値を0.02binで、Infiniumデザインごとに分類したヒストグラムを

図2：Infinium IおよびInfinium IIのケミストリーによりメチル化の全スペクトルをカバー



Infinium IおよびInfinium IIのケミストリーの相違は、特徴的なβ値分布として現れます³。

表示しています。一般的に、β値分布の両端に位置するピークが、Infinium IプローブではInfinium IIプローブよりもさらに両端に広がっており、メチル化の全スペクトルがより幅広くキャプチャーされています。

Infinium IアッセイとInfinium IIアッセイの相違は、HumanMethylation450 BeadChipによって得られるデータの精度または再現性には影響を与えません。ここで重要なのは、HumanMethylation450 BeadChipは本来、単一サンプル中でのInfinium IアッセイとInfinium IIアッセイ、またはその他の2種類のアッセイ間の比較を想定していないということです。むしろ、異なるサンプル間またはサンプル集団間での、特定のサイトの比較のために使用することが推奨されています。データ品質の測定は、このような推奨された比較に関して評価されたもので、繰り返し実験による再現性は、98%を超える相関を示し、多くの場合、99%を超えるレベルに達しています。

Infinium IアッセイとInfinium IIアッセイのケミストリーが異なることから、当社では他のメチル化評価法である全ゲノムバイサルファイトシーケンシング (WGBS) との相対的な相関関係に基づいて、Infinium IアッセイとInfinium IIアッセイ間に性能の差の有無を検討してみました。癌性肺組織および正常肺組織から単離されたDNAを、Infinium HumanMethylation450

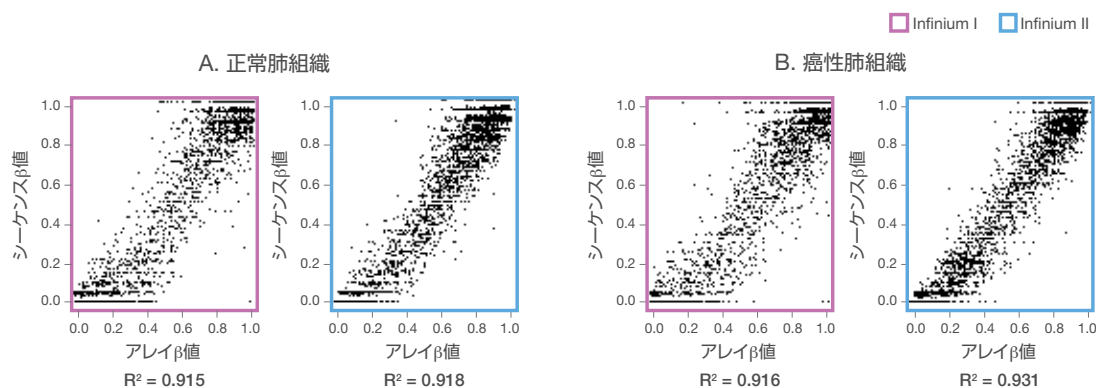
BeadChipとWGBSの両者によって評価されました (図3)。WGBSデータは、フィルターされているため20リード以上のカバレッジが必要ですが、HumanMethylation450 BeadChipでは、データの最小検出レベルは $p < 0.01$ に設定しました。Infinium IアッセイとInfinium IIアッセイを分離し、各サブセットと、WGBSによる測定に相当するサイトの間の相関レベルを測定しました。Infinium IプローブとInfinium IIプローブの、癌組織サンプルおよび正常組織サンプルにおける相関レベルは、極めて類似していました。正常組織サンプルに関しては、 r^2 値がInfinium Iプローブでは0.915、Infinium IIプローブでは0.918でした。癌性組織サンプルに関しても、結果はInfinium IプローブとInfinium IIプローブにおいて同等であり、 r^2 値はそれぞれ0.916および0.931という値を示しました。このことから、Infinium IプローブおよびInfinium IIプローブでは、そのケミストリーは相違していますが、同様な性能を提供することが、同一コントロールによる比較から示されました。

アッセイ領域におけるSNPsの存在の提示

Infinium HumanMethylation450 BeadChipのコンテンツは、メチル化研究の専門家の推奨に基づいて選択されています。コンソーシアムにより選別された領域をカバーするターゲットサイト候補は、標準的なInfiniumデザインパラメータに基づいてフィルターされています。このフィルターリングにより、プローブおよび検出サイトから、dbSNPに報告されている既知のDNA変異の位置と重複するサイトを排除します。プローブおよび検出サイトにSNPが存在することのリスク、およびそのようなリスクあるデータ解析の段階で提示する手法はすでに、GoldenGate®およびInfiniumメチル化アレイの両アッセイにて報告されています^{4,5,6}。

コンソーシアムにより選ばれた領域が、既知のSNPサイトを含まざるえない場合には、マニフェスト中にSNPサイトの存在を示す注釈が付されます。これにより、そのようなサイトを使用するか否かを、個々の研究者が決定することが可能となります。研究者は、研究対象集団の構成、SNPの検出サイトとの関係、SNPのMAFおよびその他の要因に基づいてリスクを評価し、これらのデータを利用するか否か、またどのような検証法を適用するかを独自に判断することができます。

図3：Infinium IおよびInfinium II プローブのWGBSに対する相対的相関



正常肺組織 (A) および癌性肺組織 (B) のサンプルを、HumanMethylation450 BeadChipおよびWGBSを使用して評価しました。Infinium I プローブ (紫色) とInfinium II プローブ (青色) には、高い相関性が観察されました。

図4： HumanMethylation450 BeadChipマニフェストファイルデータ

HumanMethylation450 BeadChipマニフェストファイルでは、データをSNPの位置によって以下の2つのカラムに分類します。

- (1) SNPの位置が、プローブの3'-末端の最初の10塩基以内にある。
- (2) SNPの位置が、プローブの3'-末端の最初の10塩基以降にある。

イルミナの提供する解析ツール

SNPアクセシの簡単な評価を可能とするために、イルミナではソースの異なるSNP情報として、HumanMethylation450 BeadChipマニフェストファイルおよびSupplementary SNPファイルの2種類を提供しており、いずれもMyIlluminaカスタマーポータルから入手することが可能です。HumanMethylation450 BeadChipマニフェストファイルには、既知のSNPsの存在に関する情報が含まれています(図4)。検出サイトに近いSNPほど、データに影響を与える可能性が大きいという仮定を基に、データは、SNPの位置がプローブの3'-末端から10塩基以内に存在するか、10塩基以降に存在するかによって各カラムに分類されます。

さらに包括的な、Supplementary SNPリストは、3ヶ月ごとに更新され、各アクセシに関して、プローブ配列内のSNPのリストが提供されます。Infinium IIデザインの場合にはさらに、検出サイトにおける情報も提供されます。このため、プローブの3'-末端から最初の10塩基と残りの40塩基に分類するのではなく、プローブ内での正確な位置が示されます。この情報により、研究者は独自の閾値に基づいてデータをフィルタすることが可能となります。さらに、各SNPに関して、最もレポート頻度の高いマイナーアレル頻度(MAF)が得られ、低リスク(MAF情報およびプローブのオーバーラップサイトに基づいて)のケースでも、他のソースを参考にする必要なく検出することが可能です。単一のMAFのみが示されていますが、SNPによっては、種々のMAFにおいて、複数の集団にレポートされる可能性もあることに注意してください。この情報はdbSNPから直接、容易に入手することが可能な場合もあります。Supplementary SNPリストからの情報は、GenomeStudio® Data Analysis Software Methylation Moduleに、直接取り入れることが可能です(図5)。


まとめ

HumanMethylation450 BeadChipを用いることで、1サンプルあたり、485,000を超えるメチル化サイトを1塩基の解像度で調べることが可能となります。HumanMethylation450 BeadChipの独自のデザインは、Infinium I アッセイとInfinium II アッセイのケミストリーを組み合わせ、メチル化解析のカバレッジをより広範囲なものとしします。このテクニカルノートに記載した情報およびリソースを、HumanMethylation450 BeadChipを用いて得られたデータの解析にご利用いただくことで、この強力な解析ツールを研究に最大限に活用していただくことが可能となります。

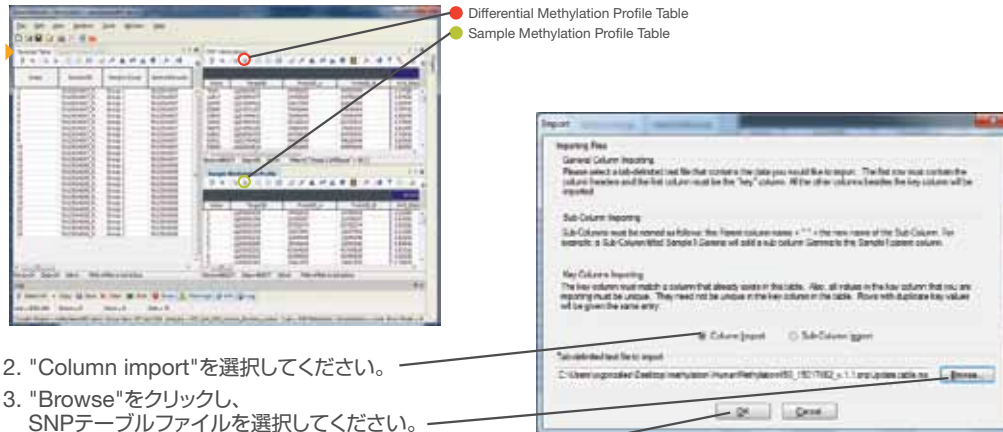
参考文献

1. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 38: 1378–1385.
2. Shoemaker R, Deng J, Wang W, and Zhang K (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.* 20: 883–889.
3. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, et al. (2011) High-density DNA methylation array with CpG resolution. *Genomics* 98: 288–295.
4. Hinoue T, Weisenberger DJ, Lange CP, Shen H, Byun HM, et al. (2011) Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res.* June 9 [Epub ahead of print].
5. Noshmeh H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, et al. (2010) Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* 17:510–512.
6. Laffaire J, Everhard S, Idhail A, Criniere E, Marie Y, et al. (2011) Methylation profiling identifies 2 groups of gliomas according to their tumorigenesis. *Neuro. Oncol.* 13: 84–98.

図5：Supplementary SNPリストのGenomeStudio Methylation Moduleへの取り込み

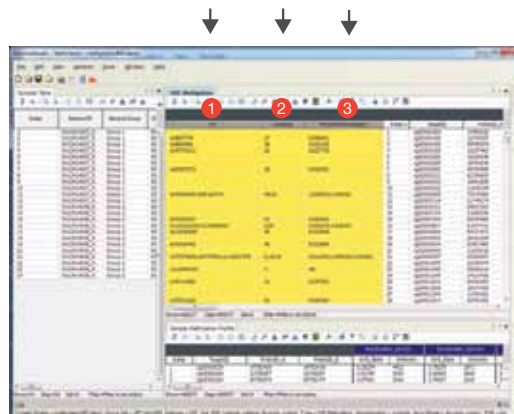
1.  をクリックして、カラムをデータベースに取り込みます(3つの選択肢があります)。

注意:Differential Methylation Profile Table、Sample Methylation Profile Table、またはGroup Methylation Profile Table(ここには示してありません)。



2. "Column import"を選択してください。
3. "Browse"をクリックし、SNPテーブルファイルを選択してください。
4. "OK"をクリックし、SNP情報を
選択したテーブルに取り込んでください。

取り込み終了後、新たに3つのカラムが表の左側に現れます。
ここでは、SNPデータをDifferential Methylation Profile Tableに取り込んだ例を示しています。



Supplementary SNPリストは、3ヶ月ごとに更新され、上記の4つのステップに従って、GenomeStudio Methylation Moduleのプロジェクトテーブルに直接取り込むことが可能です。

イルミナ株式会社

〒108-0014
東京都港区芝5-36-7 三田ベルジュビル22階
Tel (03)4578-2800 Fax (03)4578-2810
www.illumina.co.jp

代理店

本製品の使用目的は研究に限定されます。

© 2013 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSeq, DASL, DesignStudio, Eco, GAlx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, the Genetic Energy streaming bases design は Illumina, Inc. の商標または登録商標です。

その他の会社名や商品名は、各社の商標または登録商標です。予告なしに仕様を変更する場合があります。

Pub. No. 270-2012-J001 11MAY12

illumina®