

論文の内容の要旨

論文題目 次世代シーケンサーを用いたヒト大腸癌
細胞株のゲノムワイドなDNAメチル化解析
氏名 生越 克己

近年、DNAの一次構造を変化させずに遺伝形質の発現を制御、維持する機構として後天的修飾（エピジェネティクス）の概念が、生命の様々な事柄に関わる重要な因子であると認識された。エピジェネティクスの概念には、DNAのメチル化とヒストン修飾という2種類が存在する。DNAメチル化とは、DNAを構成する4種類の塩基対のうち、シトシンの5位炭素原子がDNAメチル化酵素によってメチル化されメチルシトシンとなる現象を指し、ゲノムDNAが生理的条件下で受ける唯一の化学修飾である。哺乳類に存在するシトシンのメチル化は、大部分がシトシンとグアニンが連続したCpGヌクレオチドで起こる。多くの遺伝子のプロモーター領域には、多数のCpGヌクレオチドが存在するCpG islandと呼ばれる領域が存在し、その領域のCpGヌクレオチドのメチル化は、隣接する遺伝子の発現を抑制する事で知られている。細胞のDNAメチル化は、発生分化の過程で形成されるだけでなく、細胞分裂におけるDNA複製時にも維持され、正常細胞のX染色体の不活化やゲノムインプリンティング、Alu配列などの外来性DNAの不活化を制御するメカニズムとして重要な役割を担っている。DNAメチル化の異常を引き起こす因子としては、加齢、慢性炎症、ウイルス感染、ピロリ菌感染、喫煙などがあり、これらの要因の影響を持続して受けることにより、特定領域のDNAメチル化異常が誘発される。現在、癌におけるDNAメチル化の異常が明らかにされると共に、癌の診断や予後のマーカー、治療の標的として臨床応用する試みがなされている。DNAメチル化は、癌以外の神経、免疫、代謝等の疾患においても関与する可能性が指摘され、現在最も注目されている研究分野の一つである。

近年開発された次世代シーケンサーは、超並列処理配列決定法により大量の短いDNA断片を高速に処理し、塩基配列を決定することを特徴としている。この技術は、ゲノム全体のヒストン修飾の決定、ヒトゲノムのリシーケンシ、転写開始点の検索、完全長RNAのシーケンシ、small RNAの検索、SNPs解析、転写因子の結合領域の検索等に用いられるようになり、多岐の分野で応用されるようになった。この技術はDNAメチル化を検索する方法としても使用され、MethylC-seq法、MeDIP-seq法、MSCC法などの様々な方法が開発されている。次世代シーケンサーは、シーケンシを行う事自体にかかる費用はこれまでの手法と比較すると格段に高価であるが、取得できるシーケンシが格段に増えたことにより、今までと同量のシーケンシを得るために必要な単価は安くなるという特徴を有する。次世代シーケンサーを効率よく使用する為には、“大量のシーケンシ”を“安い単価”で得ることが出来るという利点を最大限に利用した方法を考える必要がある。

そこで今回、メチル化感受性制限酵素と次世代シーケンサーを併用することを考えた。制限酵素認識サイトのDNAメチル化を特異的に測定出来るメチル化感受性制限酵素を使用することは、測定される部位を特定し、解析に必要なデータ量を最小限に留める事につながる。私はこのような簡略化したデータからゲノムワイドなDNAメチル化を検出する方法 methylation-specific digital sequencing (MSDS法)を開発した。

ゲノムワイドなDNAメチル化の変化をより簡便に測定する為に、制限酵素認識サイトの数を抑えつつ、測定可能なCpG islandの数を最大限に増加させる制限酵素の組み合わせを検討した。その結果、6塩基認識のメチル化感受性制限酵素を3種類（BssHII、EagI、SacII）選択した。ヒトゲノムには、各々72,899、90,190、66,312ヶ所の制限酵素認識サイトが存在する。この3種類を組み合わせることで、全CpG island数の75%となる21,164ヶ所のCpG islandと、13,978ヶ所の遺伝子のプロモーター領域を測定することが出来る（表1）。

表 1 使用したメチル化感受性制限酵素のリスト

制限酵素名	全制限酵素認識サイト数	制限酵素認識サイトが1ヶ所以上存在する CpG island領域の数	全 CpG island数からの割合	制限酵素認識サイトが1ヶ所以上存在する遺伝子のプロモーター領域の数
BssHII (G ^A CGCGC)	72,899	13,363	47.3%	8,787
EagI (C ^A GGCCG)	90,190	14,023	49.7%	9,089
SacII (CCGC ^A GG)	66,312	15,317	54.3%	9,793
BssHII+EagI+SacII	229,401	21,164	75.0%	13,978

制限酵素名、認識配列、制限酵素サイト数、制限酵素が1ヶ所以上存在する CpG island の数と割合、制限酵素サイトが1ヶ所以上存在する遺伝子のプロモーター領域の数を示す。上から BssHII、EagI、SacII、の情報で、最後に3種類の制限酵素を組み合わせた場合のデータを表示した。プロモーター領域は TSS±500bp とした。

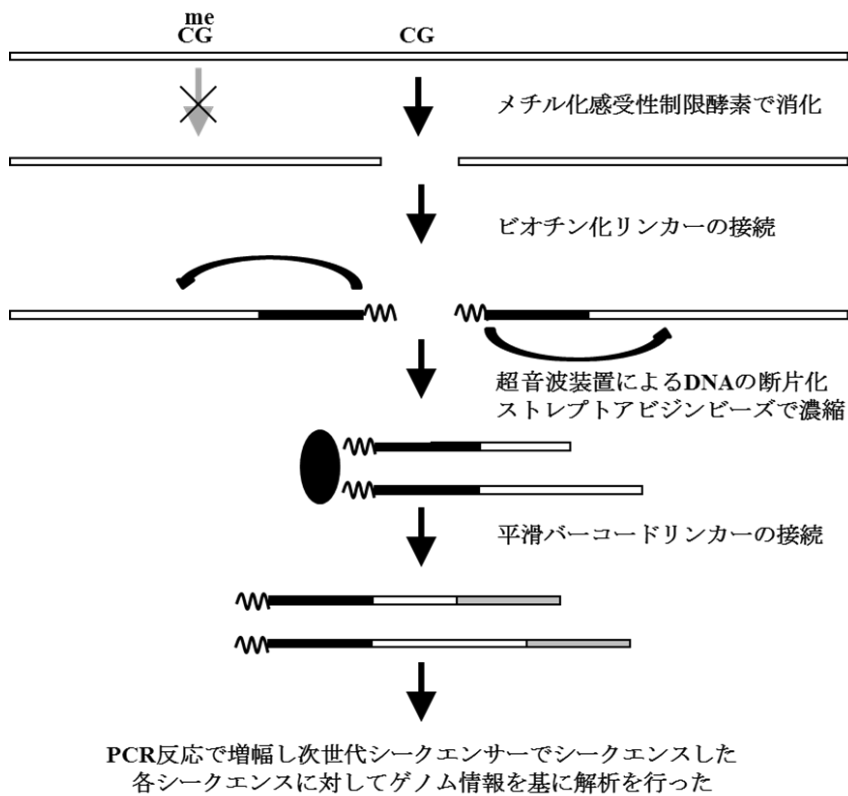


図 1 MSDS 法のサンプル作製プロセス MSDS 法では、メチル化感受性制限酵素の BssHII、EagI、SacII を使用した。メチル化感受性制限酵素は、認識サイトが DNA メチル化されていると消化されない。まず、ゲノム DNA をメチル化感受性制限酵素で消化し、その断端に、各制限酵素に対応したビオチン化リンカーを接続した。その後、超音波破碎装置で DNA を断片化し、その切断端を平滑化した。ストレプトアジジンビーズで、ビオチン化リンカーが接続した制限酵素認識サイトとその周囲の DNA 断片を回収及び濃縮した。ビオチン化リンカーの接続部位とは逆の断端に、平滑バーコードリンカーを接続した。PCR 反応で増幅した後、次世代シーケンサーでシーケンスを行い、データを得た。

メチル化感受性制限酵素は、認識配列にメチル化された CpG ヌクレオチドが存在すると消化されない。言い換えれば、メチル化されていない CpG ヌクレオチドのみが存在する制限酵素認識サイトを消化する為、その切断端に存在する認識サイトの CpG ヌクレオチドは当然メチル化されていない。MSDS 法はメチル化感受性制限酵素により切断された制限酵素認識サイトとその周囲の塩基配列を次世代シーケンサーで大量にシーケンスする方法である為 (図 1)、シーケンス tag の出現頻度の高い認識サイトとは、メチル化感受性制限酵素を使用して切断された認識サイトの事であり、その認識サイトに存在する CpG ヌクレオチドのメチル化の程度は当然低くなる。各認識サイトがシーケンスされた数=tag 数を、データ解析の過程で MSDS スコアとして扱った。

MSDS 法の有用性を確認する為、大腸癌細胞株の HT29 と HCT116 細胞に対して MSDS 法を行った。得られた tag 情報を基に、ヒトゲノムから制限酵素認識サイトを含む 100 余りの部位をランダムに選択し、その部位の DNA メチル化割合を DNA メチル化の絶対定量法である bisulfite-sequencing 法で測定し、MSDS 法を使用して得た MSDS スコアと比較した。その結果、MSDS スコアが高値であると DNA 低メチル化であるという逆相関の関係にあることが明らかとなった(図 2-A)。この結果から、制限酵素認識サイトは MSDS スコアを基に 3 種類のグループに分けることが出来た(図 2-B)。MSDS 法は同一の細胞内における DNA メチル化の有無だけでなく、異なる細胞において DNA メチル化が異なる領域の検索にも使用できる事を確認した。

次に、正常ヒト乳腺上皮細胞 (Normal human mammary epithelial cells ; HMEC) に対して MSDS 法の解析を行った。図 2-B で使用した MSDS スコアを基にしたグループ分けを、全ての制限酵素認識サイトに対して行い、ゲノムワイドな DNA メチル化を計測した。その結果、全制限酵素認識サイトは、DNA 高メチル化、低メチル化のいずれかに偏った DNA メチル化パターンを示した制限酵素認識サイトで約 90%を占めていた。遺伝

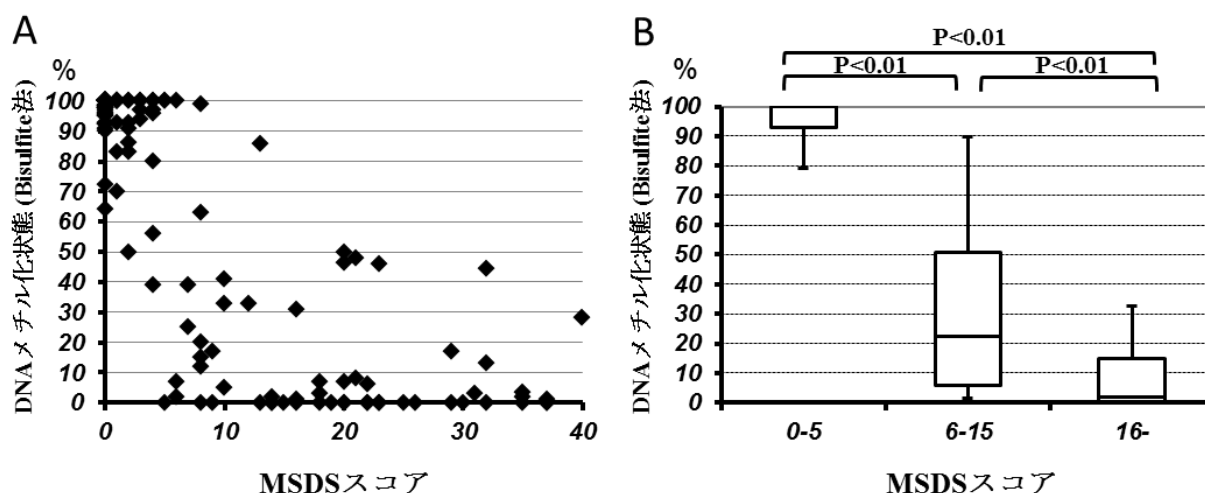


図 2 MSDS スコアと DNA メチル化 (bisulfite-sequencing 法) の関係 A. bisulfite-sequencing 法で測定した 100 ヶ所の DNA メチル化データを利用して、その部位の MSDS スコアを散布図で示した。X 軸が MSDS スコア、Y 軸が DNA メチル化状態(%)である。B. MSDS スコアを基準として、各 tag を DNA メチル化の程度によるグループ分けを行った。Kruskal-Wallis 検定を使用して解析すると、tag は 3 種類のグループに分けることが出来た。箱の下端が第一四分位値、上端が第三四分位値に対応する。線記号は中央値、ひげの下端が 5 パーセントイル、上端が 95 パーセントイルを示す。MSDS スコアが 0~5 の tag は DNA メチル化が高く(80%~100%)、MSDS スコアが 16 以上の tag は DNA メチル化が低く(0%~20%)、MSDS スコアが 6~15 の tag は中間の DNA メチル化を示した。いずれの場合も有意に差を認めた(P<0.01)。

子周囲の DNA メチル化割合は、プロモーター領域において DNA 低メチル化なサイトの割合が高く、逆に gene-body において DNA 高メチル化なサイトの割合が高いことが示された。そして、いずれの領域においても、CpG island 上のサイトはそれ以外のサイトと比較して DNA 低メチル化なサイトの割合が高いことが明らかとなった。遺伝子周囲の DNA メチル化サイトの割合とその遺伝子発現量の関係では、発現量が高い遺伝子は転写開始点周囲に DNA 低メチル化なサイトの割合が高い、逆に発現量の低い遺伝子は DNA メチル化サイトの割合が転写開始点からの距離に依存しないという結論に至った。細胞ごとに異なる DNA メチル化割合を示した領域 differentially methylated regions (DMRs)の存在部位は CpG island shore (CpG island の周囲±2000bp の範囲) に有意に多く観察されたが、遺伝子の TSS 周囲にはあまり存在しないことが示された。しかし、遺伝子発現量との関係において、プロモーター領域の DMRs と遺伝子発現量との相関を認めた。以上の結果より、プロモーター領域の DMRs は数自体少ないが、その多くが“遺伝子発現を制御する働きを持つ DMRs”であることが示唆された。

プロモーター領域に DMRs が存在して、遺伝子発現量が 5 倍以上変化していた遺伝子として、cell cycle に関与する SMPD3、VASH1、TACC1 遺伝子や、apoptosis に関与する RIPK3、TNFSF9 遺伝子、cell adhesion に関与する CLDN1、NELL1、ROBO1、ENG、PLEKHC1 等の発癌過程に重要な遺伝子が挙げられた。これらの遺伝子の中には既ががん抑制遺伝子として機能し、遺伝子発現が DNA メチル化によりコントロールされている事が報告されているものもあるが、新たに TACC1、CLDN1、PLEKHC1 の遺伝子において DNA メチル化との関係を示唆された。今回注目した遺伝子周囲のメチル化の変化は悪性度や転移のしやすさ等の癌細胞個々の特徴をとらえている可能性が考えられた。

今回の結果から、MSDS 法は遺伝子のプロモーター領域や CpG island に限らず、様々な部位の DNA メチル化を測定する事が可能であることが明らかとなった。さらに、細胞集団の DNA メチル化を比較的簡便に測定、比較を行う際に有用な方法であると考えられた。癌をはじめとした様々な疾患に関連する DNA メチル化異常を同定し、癌の診断や新たな分子標的薬剤の開発などの有用な DNA メチル化パターンの同定に役立つと考えられる。今後、DNA メチル化測定法に必要とされるサンプル量が 1 分子レベルまで減量することが可能になれば、検診のような簡便なスクリーニング検査の場で、疾患に侵される前に、疾患発症のリスクを知ることが出来る可能性がある。この事は予防医学から見るに、大いに役立つものと考えられる。MSDS 法の研究がこれからのメチル化測定のさらなる発展の一助となることを期待している。