

令和 5 年 5 月 29 日現在

機関番号：12601

研究種目：若手研究

研究期間：2021～2022

課題番号：21K15074

研究課題名（和文）微量サンプルからの長鎖全ゲノムメチル化解析手法の開発

研究課題名（英文）Development of method for long-read whole genome methylation analysis from small amount of DNA

研究代表者

関 真秀（Seki, Masahide）

東京大学・大学院新領域創成科学研究科・特任准教授

研究者番号：90749326

交付決定額（研究期間全体）：（直接経費） 3,600,000円

研究成果の概要（和文）：本研究では、塩基変換法と長いIDNAの配列を読み取ることができるナノポアシーケンスを組み合わせることで、少ないIDNAから実施できる長いIDNAのメチル化修飾解析手法nanoEMの開発を行った。nanoEMは10 ngのDNAから実施できることや臨床サンプルで実施できることを確認した。また、nanoEMで得られたメチル化データが一般的に用いられている短いIDNAを読み取るメチル化解析手法と比較しても、整合性のあるデータが取得できることを示した。さらに、既存の方法で解析が難しかったインプリンティング領域、繰り返し配列、構造変異といった領域についてもnanoEM法で解析できることを示した。

研究成果の学術的意義や社会的意義

長鎖シーケンスによるメチル化解析によって、よく用いられてきた短鎖シーケンスで解析することが難しかったゲノム領域について解析することが可能になった。しかし、長鎖のメチル化解析は、必要なDNA量が多く、臨床サンプルなどの少ない量のDNAしか取れないサンプルの解析をすることができないという問題点が存在していた。本研究で開発したnanoEM法により、これらの微量サンプルの解析が可能となった。この方法を用いることにより、様々な研究分野で今後新たな知見を生む可能性があると考えられる。

研究成果の概要（英文）：In this study, we developed nanoEM, a method for whole genome long-read DNA methylation analysis that can be performed with less DNA by combining a base conversion method with nanopore sequencing. We confirmed that nanoEM can be performed using as little as 10 ng of DNA and using clinical samples. NanoEM can obtain consistent data compared to the commonly used methylation analysis methods. We also showed that nanoEM can be used to analyze methylation patterns within regions such as imprinting regions, repetitive sequences, and structural variation, which have been difficult to analyze with most of existing methods.

研究分野：ゲノム科学

キーワード：DNAメチル化 エピジェネティクス ナノポアシーケンス

1. 研究開始当初の背景

哺乳類では、CpG 配列は高頻度でメチル化されており、DNA のメチル化が発生、がん、加齢などの様々な過程において重要な役割を持つことが知られている。DNA のメチル化によりインプリンティング制御を始めとした遺伝子発現制御やレトロトランスポゾンの不活性化が行われる。活性化したレトロトランスポゾンの転移ががんの構造変異の引き金となることが示されるなど、メチル化解析の重要性がさらに高まってきている。

これまでの DNA メチル化解析では、バイサルファイト処理により非メチル化シトシンをチミンに変換した DNA をイルミナ社などの短鎖シークエンサーで読み取る全ゲノムバイサルファイトシークエンス(WGBS)などの手法が一般的だった。バイサルファイト処理は化学的に過酷な条件下での反応であるため DNA が分解されて長い DNA が得にくく、また偏った領域がシークエンスされることが知られていた。近年、TET2/APOBEC を用いて酵素的に塩基変換を行う EMseq と呼ばれる手法が開発され、分解の少ない DNA を得られるようになった。

イルミナシークエンサーの読み取り長は、数百 bp と短い。そのため、メチル化情報は CpG 配列ごとの平均のメチル化率として扱われることが一般的であった。その読み取り長を超えて 1 本の染色体や 1 分子の DNA がどのようなメチル化パターンを有しているのかや、短いリードのアラインメントが難しいレトロトランスポゾンははじめとする反復性の高い配列や構造変異周辺のメチル化状態は、ほとんどの場合、明らかではない。近年登場したナノポアシークエンサーにより、10 kb 以上の DNA を 1 本のリードとしてシークエンスし、電子シグナルのパターンからメチル化状態を推定することが可能となった。しかし、ナノポアシークエンスのインプットとして 1 µg 程度の DNA が必要となるため、希少な細胞種や生検由来の組織などの微小なサンプルで実施することは事実上不可能である。また、長いリードを得るためには専用のキットなどを用いて傷のない DNA 調製することが重要であるが、実際は傷のない DNA を調製することが難しい場合が多い。長鎖シークエンスと塩基変換を組み合わせるメチル化を解析する SMRT-BS や lrTAP などの手法は存在するものの、特定の領域を PCR 増幅してシークエンスする手法であり、長鎖リードと塩基変換を組み合わせるゲノムワイドなメチル化解析はまだ開発されていない。また、それらの解析に用いられている解析パイプラインは、シークエンスエラーの少ない短鎖リード用のものや塩基変換を想定していない通常の長鎖リード用のものを使用しているため、十分な検出が行えないという問題点が存在する。長鎖リードのメチル化解析自体がまだ新しい技術であるため、どの組織がどのようなメチル化パターンを示すのかなどの基盤となるデータが存在していない。

2. 研究の目的

本研究は、以下の 3 つを目的として実施した。

微量サンプルからの長鎖全ゲノムメチル化解析手法 " nanoEM " の開発

微量のゲノム DNA からの塩基変換した長鎖 DNA を調製し、ハイスループット型のナノポアシークエンサー PromethION でシークエンスする nanoEM 法を確立する。

塩基変換された長鎖リードの情報解析パイプラインの整備と nanoEM の評価

塩基置換を有した長鎖リードの解析のための情報解析パイプラインの整備が不十分であるため、本研究で新たに構築して nanoEM のデータ解析に用いる。nanoEM の解析結果を既存のメチル化解析手法と比較することにより、nanoEM のメチル化解析の精度を評価する。

同一ヒト由来の多臓器の DNA への応用と長鎖メチル化解析基盤データの作出

ヒトの様々な臓器由来のゲノム DNA を使用し、長鎖メチル化解析を行う。これにより様々な組織由来のサンプルにおいても実施可能であることを確認するとともに、1 個体の臓器間で異なるメチル化領域やインプリンティング制御領域の検出を試みる。

3. 研究の方法

微量サンプルからの長鎖全ゲノムメチル化解析手法 " nanoEM " の開発

メチル化状態の異なる乳がん細胞株 MB231 と BT474 を用いて、nanoEM 条件検討を行った。DNA 量については、1-50 ng のゲノム DNA を用いた。EMseq により塩基変換を行った後、ロング PCR により変換 DNA の全ゲノム増幅を行った。得られた DNA を PromethION によりシークエンスした。

塩基変換された長鎖リードの情報解析パイプラインの整備と nanoEM の評価

ショートリードのメチル化解析に用いられる Bismark の解析戦略をロングリード用のアラインメントソフトウェア minimap2 に応用することで、長鎖リード用のメチル化解析パイプラインを開発した。比較データとして、ショートリードの WGBS、EMseq 及び、nanopore の全ゲノムシークエンスからメチル化検出を実施して、nanoEM データの評価に用いた。

同一ヒト由来の多臓器の DNA への応用と長鎖メチル化解析基盤データの作出

2 検体の多臓器 (6 および 7 臓器) について、PromethION による全ゲノムシークエンス (WGS) を行った。それぞれ検体の各 1 臓器については、短鎖シークエンスによる WGS も行った。

4. 研究成果

微量サンプルからの長鎖全ゲノムメチル化解析手法 "nanoEM" の開発

上記 2 種類の細胞株を用いて nanoEM 法の開発を行った。DNA を 10 kb 弱に断片化後、EMseq キットを使用してアダプター付加や塩基変換を行った。変換後の長鎖 DNA を増幅するために、複数の PCR 酵素及び条件を検討した。その結果、変換された長鎖 DNA の増幅に成功し、nanoEM のプロトコルを確立することができた(図 1A)。nanoEM 法を用いて、1-50 ng の DNA で検討した結果、いずれの DNA 量からでも nanoEM のライブラリーを得ることができた。PromethION でシーケンスすることにより、塩基変換した 5 kb 程度の長さのリードを取得できることを確認した(図 1B 及び C)。また、ヒト乳がん 3 症例(正常部とがん部の 3 ペア)由来の DNA を用いて nanoEM ライブラリーを調製し、臨床検体においても nanoEM が実施可能であることを示した。

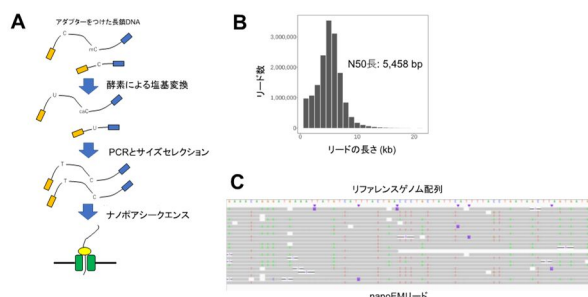


図 1. nanoEM の概要

(A) nanoEM 法の概略図。(B) nanoEM のリー長の分布。(C) nanoEM リードのゲノムへのマッピング結果の例。赤と黄色が T と A への塩基置換を示している。

塩基変換された長鎖リードの情報解析パイプラインの整備と nanoEM の評価

nanoEM データの解析のためのデータ解析パイプラインを開発し、github に公開した(<https://github.com/yos-sk/nanoEM>)。開発したパイプラインを用いて、nanoEM データを解析した。他の既存法により取得されたメチル化情報と比較したところ、10 ng 以上の DNA からであれば nanoEM は既存の手法と非常に高い相関を示すデータが取得できることが確認された。nanoEM では、短鎖リードで解析が困難であった染色体ごとにメチル化状態が異なるインプリンティング領域(図 2A)、繰り返し配列、構造変異周辺のメチル化状態(図 2B)を検出できることを示した。

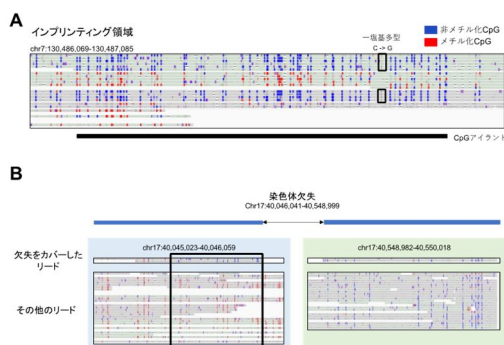


図 2. インプリンティング領域と構造変異の解析結果

(A,B) nanoEM リードのインプリンティング領域(B)と SV(B)へのマッピング結果。(A) 四角で染色体特異的な SNP を示した。(B) 四角で示した領域で欠失を持つ DNA のみに特異的な低メチル化状態が観察された。

DNA メチル化解析においては、全ゲノムを対象とした解析よりも CpG アイランドなどのメチル化による制御を受ける領域を高深度に解析することが重要である。そこで、Exome シークエンスなどに使用されるハイブリキャプチャー法を nanoEM と組み合わせることで、調節領域特異的な長鎖メチル化解析法 targeted nanoEM 法(t-nanoEM)の確立を試みた。まず、塩基変換前の DNA をキャプチャーする方法で検討を行ったところ、長い DNA が得られないという問題点が判明した。そこで、塩基変換及び PCR 増幅後の DNA をキャプチャーする方法に切り替えたところ、5 kb を超える長い DNA が取得できた。また、微量化の検討を行ったところ、10 ng の DNA から実施することができることを確認した。現在、さらなる改良を実施しており、乳がん細胞株及び臨床検体において解析を行った後に論文投稿を予定している。

同一ヒト由来の多臓器の DNA への応用と長鎖メチル化解析基盤データの作出

ヒト 2 検体について、短鎖シーケンスのデータから、SNP 検出を行った。また、各臓器のナノポアシーケンスデータから nanopolish によるメチル化検出を行った。Whatsap を用いて、参照の SNP 情報と長鎖リードを解析することで、染色体ごとに長鎖リードの振り分けを行った。振り分けられたリードのメチル化情報を集計することで、染色体ごとにメチル化状態の異なる領域を 400-800 箇所程度検出することができた。現在、検体間の同一組織での共通性の有無や臓器特異的に染色体ごとにメチル化状態の異なる領域などに着目した解析を進めている。

業績 1 と 2 の研究成果については、国内外の学術雑誌への論文発表(Sakamoto et al. *Nucleic acid research* 2021; Zaha et al. *Bio-protocol* 2022; 関他. *月刊細胞* 2021)、学会発表(関他 第 44 回分子生物学会年会 2021; Seki & Suzuki. *AGBT* 2023)、及び、招待講演(関. 第 15 回日本エピジェネティクス研究会 2022; 関、鈴木. 第 8 回日本 HTLV-1 学会学術集会 2022)を行った。また、所属研究科からプレスリリースを行った(<https://www.k.u-tokyo.ac.jp/information/category/press/8599.html>)。3 の成果についても、解析を進めており、結果がまとまり次第、論文投稿を予定している。また、実施過程で得られたロングリードの情報解析の知識を活かして書籍の分担執筆(三宅、他. *がんゲノムデータ解析* 2022)を行った。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 2件）

1. 著者名 Sakamoto Yoshitaka, Zaha Suzuko, Nagasawa Sato, Miyake Shuhei, Kojima Yasuyuki, Suzuki Ayako, Suzuki Yutaka, Seki Masahide	4. 巻 49
2. 論文標題 Long-read whole-genome methylation patterning using enzymatic base conversion and nanopore sequencing	5. 発行年 2021年
3. 雑誌名 Nucleic Acids Research	6. 最初と最後の頁 e81
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nar/gkab397	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Suzuko Zaha, Sakamoto Yoshitaka, Nagasawa Sato, Sugano Sumio, Suzuki Ayako, Suzuki Yutaka, Seki Masahide	4. 巻 12
2. 論文標題 Whole-genome Methylation Analysis of APOBEC Enzyme-converted DNA (~5 kb) by Nanopore Sequencing	5. 発行年 2022年
3. 雑誌名 BIO-PROTOCOL	6. 最初と最後の頁 e4345
掲載論文のDOI（デジタルオブジェクト識別子） 10.21769/BioProtoc.4345	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 関 真秀、坂本 祥駿、座波 紗子、鈴木 穰	4. 巻 53
2. 論文標題 ナノポアシーケンサーによるDNAメチル化解析	5. 発行年 2021年
3. 雑誌名 月刊「細胞」	6. 最初と最後の頁 818-820
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 2件 / うち国際学会 1件）

1. 発表者名 関 真秀、坂本 祥駿、座波 紗子、永澤 慧、鈴木 絢子、鈴木 穰
2. 発表標題 微量DNAからの長鎖DNAのメチル化解析手法の開発
3. 学会等名 第44回日本分子生物学会年会
4. 発表年 2021年

1. 発表者名 関 真秀
2. 発表標題 微量DNAからの長鎖DNAメチル化解析手法の開発
3. 学会等名 第15回日本エピジェネティクス研究会年会（招待講演）
4. 発表年 2022年

1. 発表者名 関 真秀、鈴木 穣
2. 発表標題 次世代シーケンス技術の現状と展望
3. 学会等名 第8回日本HTLV-1学会学術集会（招待講演）
4. 発表年 2022年

1. 発表者名 Masahide Seki, Yutaka Suzuki
2. 発表標題 Targeted long-read methylation sequencing using nanopore sequencing with hybridization capture.
3. 学会等名 AGBT 2023（国際学会）
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

少量のDNAから実施できる長いDNAの全ゲノムメチル化解析法を開発
<https://www.k.u-tokyo.ac.jp/information/category/press/4024.html>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------