

日本医療研究開発機構 創薬等ライフサイエンス研究支援基盤事業 事後評価報告書



I 基本情報

補助事業課題名：（日本語）創薬等ライフサイエンス研究支援基盤事業
（プログラム名）（英語）Platform Project for Supporting Drug Discovery and Life Science Research

実施期間：平成29年4月1日～令和4年3月31日

補助事業担当者 氏名：（日本語）伊藤 隆司
（英語）Takashi Ito

補助事業担当者 所属機関・部署・役職：
（日本語）国立大学法人九州大学・大学院医学研究院・教授
（英語）National University Corporation Kyushu University, Faculty of Medical Sciences, Professor

II 補助事業の概要

本事業は、我々が独自に開発した世界最高感度・精度の全ゲノムバイサルファイトシーケンス技術である Post-Bisulfite Adaptor Tagging (PBAT)法の高度化を進めること、更にそうして高度化されたPBAT法を基盤とする先進的なメチローム解析支援を幅広い層の研究者に提供することを目的とする。そのために必要な要素技術および周辺技術の開発のみならず、支援希望者の要望に沿った有用な応用展開も実施した。

1) PBAT法の高度化とそれを用いたメチローム解析支援

PBAT法は、バイサルファイト変換反応後の変性1本鎖DNA(ssDNA)を鋳型に2回のランダムプライミングによって全ゲノムバイサルファイトシーケンシング(WGBS)ライブラリを調製する技術であり、微量試料からの一塩基解像度メチローム解析に突破口を開いた。PBATは一細胞解析にも利用されているのみならず、最も正確かつバイアスが少ないことも報告されるなど、世界標準のWGBS技術としての地位を確立した。しかしながら、PBATには、①ライブラリ断片長が短いこと、②極微量解析時のマッピング率が低いこと、③GC含量の低い領域のカバレッジが悪いこと、という3つの弱点が残されていた。これらはいずれもランダムプライミングに起因する。

そこで我々は、ランダムプライミングに代わる新しいssDNAへのアダプタ付加法の開発に取り組んだ。その結果、独自のssDNA連結技術TACSライゲーション(TdT-assisted Adenylate Connector-mediated ssDNA ligation)

を開発した。TACS ライゲーションは、ssDNA の 3'末端に TdT (Terminal deoxyribonucleotidyl Transferase) を利用して 2-3 残基のアデニル酸を付加した後に、TS2126 RNA リガーゼを用いて ssDNA アダプタと高い効率で連結する技術である。つまり、TdT による限定的リボテーリングによって付加されたアデニル酸を RNA リガーゼによる効率的連結のためのコネクタとして利用するのが、TACS ライゲーションであり、画期的な効率で ssDNA 同士の連結を実現できる。

こうして開発した TACS ライゲーションを PBAT に導入するために、まず我々は PBAT の 2 回目のランダムプライミングを TACS ライゲーションで置換する tPBAT の開発に取り組んだ。その結果、上記の欠点①と②の克服に成功して、データ量およびデータ品質の改善を達成した。tPBAT は直ちに支援に投入され、すべての支援課題において所期の WGBS データ取得を達成し、その優れた実用性が実証された。

一方、tPBAT では上記の欠点③は解消されなかった。その理由は 1 回目のランダムプライミングが残っているからである。これを残さざるを得なかったのは、バイサルファイト処理 DNA に対して TACS ライゲーションが有効に働かなかったからである。この事実は、バイサルファイト変換反応で生じた DNA の切断端が TdT の基質にならない構造をしていることを示唆する。バイサルファイト変換 DNA に対する脱リン酸化酵素処理が効果を発揮しなかったことから、切断端は水酸基でもリン酸基でもない構造であると考えられた。そのような構造を除去する可能性を持つ酵素を検討し、それを発現精製してバイサルファイト変換 DNA に作用させたところ、TdT による 3'末端標識効率が各段に上昇した。つまり、この末端修復反応によって、バイサルファイト変換 DNA に直接 TACS ライゲーションでアダプタを付加することが可能になった。

そこで末端修復処理後に TACS ライゲーションによって付加したアダプタに対する相補的プライマーを用いて、バイサルファイト変換 DNA の相補鎖を合成して、2 つ目のアダプタを付加する方法に検討を加えた。2 つ目のアダプタ付加に再度 TACS ライゲーションを用いる TACS-TACS 法では、欠点③が克服されゲノム全体を極めて均一性高くカバーすることが可能になった。しかしながら、TACS-TACS 法はライブラリ収量が tPBAT に及ばず、またライブラリ断片長の更なる伸長効果も弱かった。そこで相補鎖合成に用いる DNA ポリメラーゼの最適化を進めると同時に、2 つ目のアダプタ付加に T4 DNA リガーゼを用いる TACS-T4 法の開発も行なった。

2) TACS ライゲーションの応用

TACS ライゲーションは ssDNA 連結法としては、既存手法を遙かに凌駕する効率を示した。そこで PBAT 以外の様々な用途への応用を検討することとした。血中セルフリー DNA (cfDNA) は非侵襲性のバイオマーカーとして注目されており、支援コンサルティングでも cfDNA 解析を希望されることがあった。興味深いことに cfDNA 解析時に ssDNA ライブラリ作成法を利用すると、通常の解析対象であるヌクレオソームサイズの DNA 断片よりも短い断片が捕捉され、それらが転写因子結合部位などの遺伝子発現制御領域の情報をもたらしてくれることが報告された。そこで我々は TACS ライゲーションを cfDNA 解析に応用することとした。それに際して cfDNA の全貌を把握するために各種条件を検討する過程で我々は、血中には約 50 塩基長の ssDNA がヌクレオソームサイズ断片と同程度の分子数で存在しているにも関わらず、ポピュラーな cfDNA 単離法では回収されないために見落とされてきたことを見出した。この DNA を cell-free short single-stranded (3S) DNA、略して C3D と命名し、C3D も定量的に回収できる方法で調製した cfDNA を上記の TACS-T4 法でライブラリ化して解析を行った。C3D は、ゲノムにマッピングするとピークを形成する集団 C3D^{on} と形成しない集団 C3D^{off} に大別されることが判明した。後者は、以前に報告されていた転写因子結合部位などの制御領域に相当していた。一方、前者は G-quadruplex (G4) 構造の相補鎖(i-motif)に富むという特徴的な性質を備えた全く新しいクラスの cfDNA であることが判明した。今後、その生成機構や病態生理学的意義に興味を持たれる。

3) FFPE 試料からの PBAT

支援コンサルティングではホルマリン固定パラフィン包埋 (FFPE) 試料からの解析を希望する声を少なくなかった。FFPE 試料からの DNA 回収は量的にも質的にも満足ゆくものでなく、それが様々な解析の足枷となっていた。我々は、一番の問題が脱クロスリンクのステップにあると考えて、マウス肝の FFPE 標本をモデルにホルム

アルデヒドスカベンジャーである tris(hydroxymethyl)aminomethane の濃度・pH・反応時間等が、DNA 回収量・回収 DNA 断片長・PCR 効率等に与える影響を系統的に検討した。その結果、高濃度 Tris を適切な条件で用いる HiTE 法を開発するに至った。

HiTE 法を実際の病理標本（様々な年数を経たヒトの郭清リンパ節標本）に適用したところ、試験した市販キット中で最も効率のよかったもの（以下、従来法）と比較して、回収量が3倍に向上することが判明した。更に、HiTE 法と従来法で回収された DNA を同一量用いて、全ゲノムシーケンシングライブラリを作成したところ、前者の調製高率は後者のそれよりも約3倍高いことが分かった。したがって、HiTE 法を用いることによって、単位 FFPE 試料あたり約10倍量の NGS ライブラリを調製できることが分かった。最後に、HiTE 法で調製した DNA から作成したライブラリをシーケンスしたところ、既知変異の検出性能は従来法と同等であることが確認された。また、酵母核を固定したサンプルによる実験では、従来法よりもゲノムをより均一にカバーできることも判明した。以上より、HiTE 法は FFPE 試料の解析を質量の両面で改善する上で非常に有効な方法であることが判明した。

そこで HiTE 法のメチローム解析における有用性を検討するために、HiTE 法で調製した DNA から tPBAT による WGBS ライブラリ調製を行ったところ、非固定組織から回収した DNA から調製した場合と同等の収量を得ることに成功し、WGBS への応用も可能であることが示された。

The primary goal of this project is to improve post-bisulfite adaptor tagging (PBAT), which we developed as the most sensitive and accurate whole-genome bisulfite sequencing (WGBS) method, and to provide BINDS users with more advanced methylome analysis. We also pursued related technologies and their applications, according to requests from BINDS users.

1) Improved PBAT and its application to methylome analysis support

PBAT uses two rounds of random priming on denatured bisulfite-converted DNA (BS-DNA) to generate WGBS library. PBAT has not only realized single-cell methylome analysis but demonstrated its unsurpassed accuracy. Nevertheless, PBAT has three drawbacks, namely 1) short insert size, 2) low mapping rate at ultra-low input analysis, and 3) compromised coverage of GC-poor regions. As all these drawbacks likely stem from the random priming steps, we intended to develop a novel adaptor tagging method for single-stranded DNA (ssDNA) to enable random priming-free PBAT. Consequently, we developed TACS ligation (TdT-assisted Adenylate Conector-mediated sssDNA ligation), in which TdT first attaches a couple of adenylates to the 3'-end of target ssDNA and then TS2126 RNA ligase connects ssDNA adaptor to the attached adenylates with high efficiency. By replacing the second random priming step of PBAT with TACS ligation, we developed tPBAT and successfully overcame the drawbacks 1) and 2). We thus introduced tPBAT to the support of methylome analysis and succeeded in all the projects. For further improvement, the drawback 3) should be overcome via elimination of the first random priming step by directly ligating the first ssDNA adaptor to BS-DNA. For this purpose, we developed an enzymatic method to end-polish the 3'-end of BS-cleaved DNA so that it can accept TACS ligation. Indeed, WGBS library prepared from end-polished BS-DNA via two rounds of TACS ligation achieved an unsurpassed evenness in terms of genome coverage albeit a compromised efficiency in library yield compared to tPBAT. We also develop a TACS-T4 method, which uses T4 DNA ligase to attach the second double-stranded adaptor.

2) Application of TACS ligation to cell-free DNA analysis

We applied TACS ligation to cell-free DNA (cfDNA) analysis because ssDNA library preparation was reported to uncover short cfDNA fragments derived from transcription factor-binding sites. During the

course of this study, we unexpectedly revealed that human plasma contains a plenty of ssDNAs of ~50 nucleotides, which cannot be recovered by popular column-based DNA isolation methods. We termed these cell-free short ssDNA (3S) DNA as C3D and applied TACS-T4 to sequence cfDNAs isolated using a method that does not miss C3D. Subsequent data analyses revealed that C3D is composed of two subgroups, C3D^{off} and C3D^{on}. The former does not form peaks upon genome mapping and corresponds to the previously reported fraction enriched with regulatory regions, whereas the latter forms peaks and represents a totally novel cfDNA class enriched with the antisense strand of G-quadruplex (G4) structure. Biogenesis and pathophysiological significance of the latter is of particular interest in the future.

3) Methylome analysis of FFPE samples

Based on the requests from the users, we tried to improve DNA isolation from Formalin-Fixed Paraffin-Embedded (FFPE) samples. We developed a novel method termed HiTE, which improves both quantity and quality of recovered DNA. The amount of whole-genome sequencing (WGS) library DNA prepared from a single FFPE slice with HiTE is 10-fold larger than those with conventional methods. We successfully applied tPBAT to HiTE-prepared DNA and enabled WGBS from FFPE samples. HiTE will thus accelerate WGS/WGBS from precious FFPE samples.