

博士論文審査報告書

論文題目

Signal Detection and Biological
Feature Extraction for
High-throughput Data of
N6-methyladenosine (m6A)

N6-メチルアデノシン (m6A) の
ハイスループットデータの
信号検出と生物学的特徴抽出

申請者

| | |
|--------|-------|
| Yiqian | ZHANG |
| 張 | 怡倩 |

電気・情報生命専攻 バイオインフォマティクス研究

2020年10月

RNA 修飾は、RNA 配列を構成する塩基やリボースの化学修飾であり、遺伝子の発現制御などの機能に深く関わっていることが知られている。現在までに、150 種類を超える RNA 修飾が発見されている。近年、新型シーケンサーを応用した実験・測定手法により、全転写産物（トランスクリプトーム）中のさまざまな種類の RNA 修飾を、網羅的に同定することが可能となってきた。その結果、RNA 修飾の生物学的意義を明らかにすることを目的とする「エピトランスクリプトーム」と呼ばれる一つの研究分野が形成され、世界中で活発に研究がなされている。

本論文の研究対象である N6-メチルアデノシン(以下では m6A と略す)は、転写産物中にもっとも数多く見られる RNA 修飾の一つであり、メッセンジャー RNA やノンコーディング RNA などの機能を担う「機能エレメント」として、エピトランスクリプトーム研究において特に注目されているものである。網羅的実験情報に基づき m6A 修飾の全貌を明らかにするためには、情報科学の積極的な活用が必須となるが、現時点において m6A 修飾に関わる情報科学的な研究は限られており、これが研究のボトルネックとなっている。そこで、本博士論文では、m6A 修飾研究の加速および全体像の解明を目的とし、網羅的実験情報と情報科学を融合させた研究を遂行した。特に、本研究では、多くの研究者が興味を持つ 2 つの課題に焦点を当てている。一つ目が m6A 修飾やそのシグナルを同定することであり、2 つ目が m6A 修飾に関わる生物学的特徴の抽出である。

本論文は 5 章から構成される。以下に各章の詳細について記載する。

1 章では、本博士論文の研究背景について述べられている。m6A 修飾の生物学的な重要性とともに、トランスクリプトームワイドに m6A を同定する実験技術として MeRIP-seq と miCLIP について図を用いて詳しく説明がなされている。また、m6A 修飾に関連する、現在判明しているさまざまな制御因子についても説明がされている。さらに、m6A 修飾のデータ解析および情報解析に関する方法についてもサーベイがなされている。

2 章について詳細を述べる。2 章では、深層学習技術を用いて m6A を含む配列の予測とその特徴付けを行う手法 (DeepM6ASeq) について述べられている。配列から m6A 修飾を予測する既存の手法はいくつか存在するが、それらの既存手法では k-mer (連続する長さ k の部分文字列) を特徴量として用いており、解釈性に乏しいという問題があった。そこで本研究では、この問題を解決することを目的として研究を行った。提案手法 DeepM6ASeq は、畳み込みニューラルネットワーク (CNN) と長・短期記憶 (LSTM) を組み合わせた方法となっており、m6A 修飾を含む配列の予測を行うことができると同時に、予測に寄与する配列特徴 (配列モチーフ) を導出することが可能となっている。注目すべき点として、DeepM6ASeq により同定した配列モチーフから、最近論文で報告された m6A リーダーである FMR1 を同定することに成功した。さらに、顕著性マップ (saliency map) を用いることに

より，配列中の m6A 修飾の部位をその信頼度に応じて可視化することに成功した．

3章では，MeRIP-seq から得られる実験情報から m6A 修飾のシグナル領域を同定するための統計的な手法の開発を行った．MeRIP-seq は比較的安価に行える実験手法であり，m6A 修飾研究において広く用いられている．MeRIP-seq のデータ解析においては，シーケンサーにより得られるリードデータから m6A 修飾のシグナルを検出することが主要な問題となるが，既存のツールは多大な計算時間を要し，さらに，リード配列のストランド情報などが十分活用されていないなどの問題点があった．そこで本研究では，これらの問題点を解決することを目的とし，負の 2 項分布に基づいて，MeRIP-seq 実験から得られるリード情報に基づき m6A 修飾シグナルを効率的かつ高精度に同定するためのソフトウェア MoAIMS を開発した．MoAIMS は既存のソフトウェアに比べて，精度はほぼ維持したまま大幅に処理速度の向上が実現した．さらに，MoAIMS により推定された m6A シグナルは，m6A メチル化酵素をノックダウンしたデータセットで減少傾向となることが観測された．したがって，MoAIMS によって推定される m6A シグナルのバックグラウンドに対する割合は，トリートメントの影響を見る直感的な指標となっていることが示唆され，m6A 制御因子のノックダウン実験がうまく機能しているかを判定するひとつの指標としても活用可能である．

4章では，m6A 修飾の分子メカニズムに関わる特徴を抽出するために，m6A に結合する RNA 結合タンパク質 (RNA Binding Protein, RBP) を同定する情報科学のフレームワークを構築した．既存の m6A の生物学的特徴を抽出するツールは，(2章で申請者が提案した DeepM6ASeq を含め) 配列ベースのものがほとんどである．このようなツールでは，m6A 修飾に関わる配列特徴を抽出することが可能であるが，抽出された配列特徴のみから m6A 修飾の分子生物学的な意義およびメカニズムを明らかにすることは困難である．そこで本章では，m6A に結合する RBP を同定することにより，m6A の分子メカニズムの解明を目指した．本章においては，第一に，RBP 結合の網羅的実験データ (CLIP-seq) およびエンリッチメント解析により，m6A 修飾に関与するタンパク質群の同定を行った．エンリッチメント解析では，YTH ドメインを含む既知の m6A の制御タンパク質が同定された．さらに，同定したタンパク質群の結合プロファイルを用いて，機械学習 (ランダムフォレスト) による分類モデルの構築を行った結果，高精度で m6A 修飾部位の推定が行えることが示された．すなわち，RBP 結合プロファイルは m6A 修飾予測に寄与することが示され，さまざまな RBP が m6A に関連する制御に関わっていることが示唆された．

5章では，本論文のまとめと将来展望について述べられている．将来展望としては，MeRIP-seq データから m6A シグナル同定の際の統計モデルの改善，生物学的な特徴抽出手法の改善，および，N1-メチルアデノシン (m1A)

や 5-メチルシトシン (m5C) などの他の種類の RNA 修飾への応用が挙げられている。

予備審査会および公聴会においては、本論文で利用している深層学習モデルおよび統計モデルの妥当性や、評価方法などに関して多角的な観点から審査がなされた。また、将来研究についても審査が行われた。結果、本博士論文は、これらの審査プロセスを通して、4回の大幅な改訂が行われている。

以上、申請者は、重要な RNA 修飾である m6A に関して、配列からの予測と配列特徴の抽出手法の開発、大規模実験データからの m6A シグナルを同定する手法の開発、RNA 結合タンパク質に基づく m6A 修飾の生物学的な意味付けを行う手法の開発を行い、その有効性を示した。いずれの手法もソフトウェアとして実装を行い、無償で公開を行っている。これらの技術とソフトウェアは、今後の m6A 修飾研究に大きな寄与をもたらし、いまだにその多くの機能が未知である長鎖ノンコーディング RNA などの機能性 RNA の機能の解明にも資することが期待される。したがって、本論文は、博士(工学)の学位論文として十分価値のあるものであると判断できる。

2020年9月

審査員

主査 早稲田大学教授 博士(理学) 東京工業大学 浜田 道昭

年 月 日

早稲田大学教授 博士(工学) 東京大学 村田 昇

年 月 日

早稲田大学教授 博士(医学) 京都大学 井上 真郷

年 月 日

東京大学教授 Ph.D. ボストン大学 フリス マーティン

年 月 日
