

新型コロナウイルス感染症対応HPCI臨時課題
課題番号hp200145

COVID19 ナノポアシーケンスデータを用いた RNA塩基修飾の詳細解析



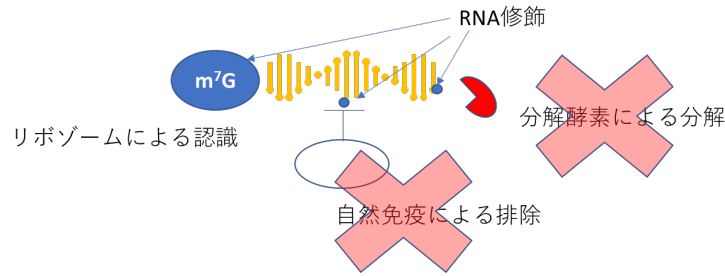
東京大学
先端科学技術研究センター

生命データサイエンス分野 講師
上田 宏生



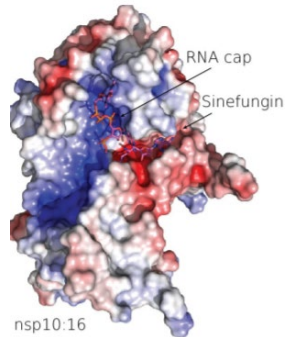
FIGHT Against COVID-19

1. RNA修飾により細胞の防御機構がすり抜けられる。



2. RNA修飾はSARS-COV2の非構造タンパク nsp10, nsp14, nsp14, nsp16により付加されることが分かっている。

- Nsp10
- Nsp13 a bifunctional RNA/NTP triphosphatase (TPase) and helicase
- Nsp14 a bifunctional 3'→5' mismatch exonuclease and mRNA cap guanine-N7 methyltransferase
- Nsp16 a ribose 2'-O methyltransferase; and an elusive guanylyl transferase



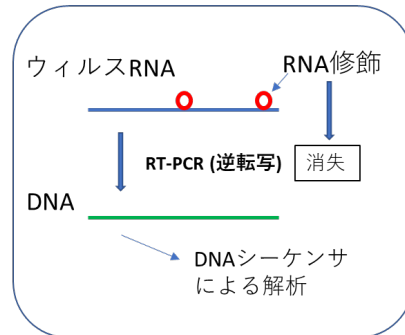
ゲノムRNAの配列解析が進んでいるが、RNAは修飾を伴って機能を持つため、RNA修飾の解析も必要である。

3. RNA修飾の解析法

既存のDNAシーケンサでは検出できず、質量分析器では位置の特定が難しい。ナノポアシーケンサでは、原理的に検出可能であるが、データ解析が課題となる。



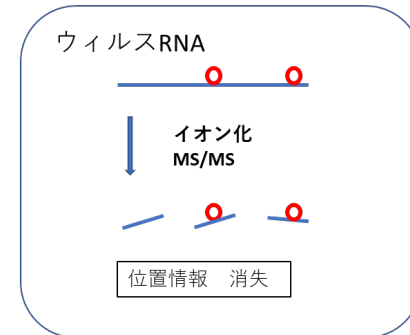
RT-PCR + 通常のシーケンサ



RNA修飾の情報はRT-PCRの過程で失われるため、通常のシーケンサは使えない



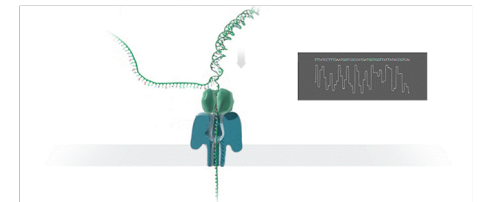
質量分析器



存在するRNA修飾の種類を特定することは可能であるが、位置情報を特定することは困難



ナノポアシーケンサ

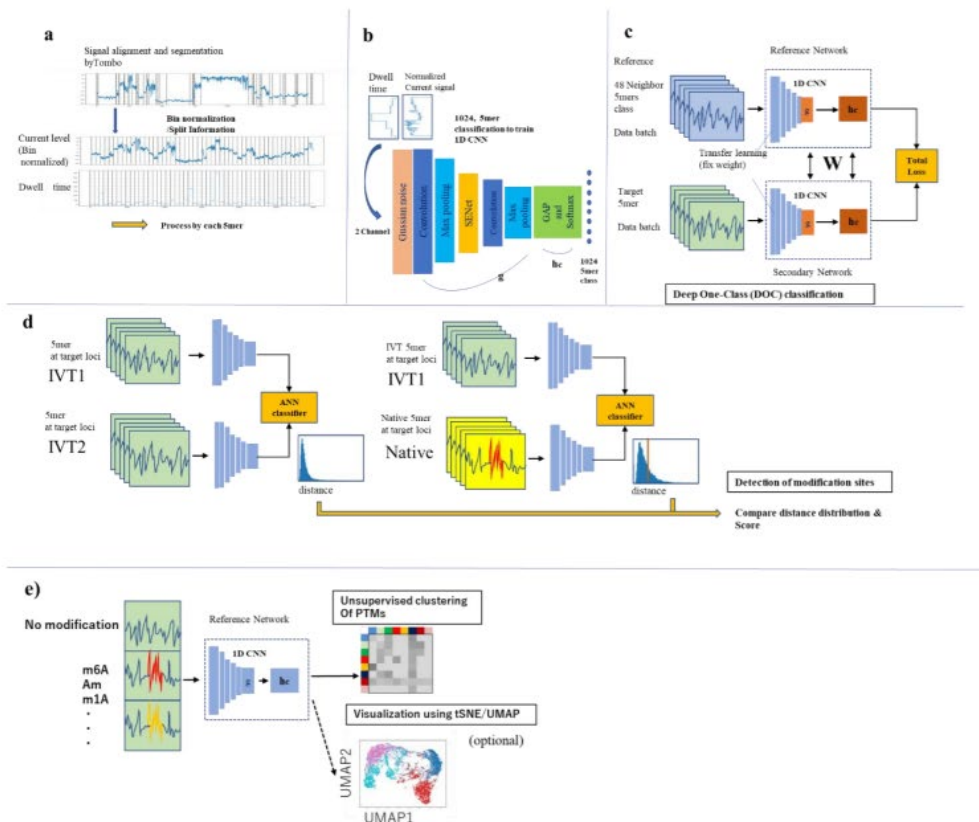


RNA 1分子ごとに解析が可能。修飾塩基は異なる電流値として検出される。



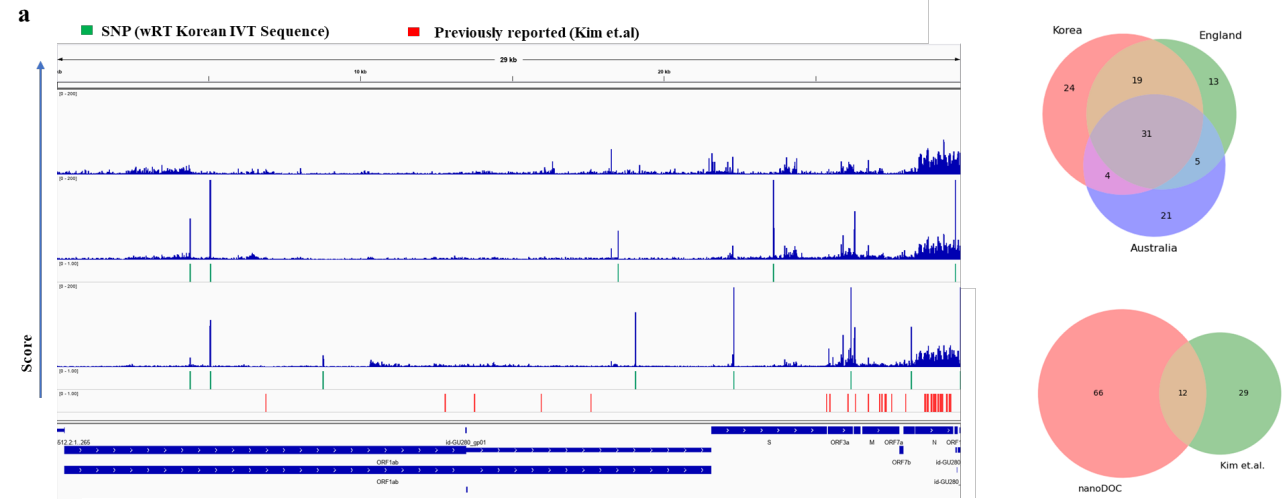
複雑な大量データの解析が課題となる

4. RNA修飾検出ソフトウェアの開発



深層学習の一手法（Deep One Class + 1D CNN）を用いて、RNA修飾解析ソフトウェアを開発し、公開した。
（下記URL）既知の方法より高い精度、感度を達成するとともに、(AUC 0.89⇒0.96)
異なる2-3種のRNA修飾の検出能力を確認した。

5. SARS-COV2 RNA修飾の検出



海外3グループのSARS-COV2データを新規ソフトウェアを用いて解析。

3つのグループで多くの修飾部位は保存されていることを確認（右上図）した。既知の方法で検出された修飾部位とは一致が少なく（右下図）、既存の方法の精度の問題が懸念される。検出部位の多くの修飾部位はNたんぱくコード上の3'末端に集中している他、全端的にまばらなピークが確認される。

3'末端上の修飾はRNaseによる分解からSARS-COV2ゲノムを守る役割があることが推測される。

今後、解析を進めウィルスゲノムの違いによる修飾の変化や感染力への影響について解析を進めたいと考えている。