

223 がんの細胞間不均一性を発見するリードカバレッジ解析	尾崎 遼
--------------------------------------	------

【目的】 がんの多くは遺伝的変異により引き起こされるが、これまでバルク RNA-seq (非 1 細胞解像度) を用いた研究から、遺伝子発現量の変動に加え、スプライシング異常や新規転写単位など多様な RNA プロセシングイベントが、浸潤性や転移性といったがんの性質の変化、ステージ判別、予後予測に重要であることが示されている。最近、1 細胞 RNA-seq が、遺伝子発現量に基づいて腫瘍内・腫瘍間不均一性を調べる手段となりつつある。しかし、既存のデータ解析ツールでは遺伝子発現量のみを用いるため、遺伝子発現量に直接現れない異常スプライシングなど RNA レベルの現象が腫瘍や周辺細胞に存在しても見逃してしまうという問題があった。この問題の解決策になるのがリードカバレッジ解析である。リードカバレッジはゲノム上の各領域から転写された RNA 量を表すシグナルの分布であり、多様な RNA プロセシングイベントを反映する。そのため、リードカバレッジを 1 細胞レベルで解析することで、がんの進行や腫瘍不均一性に関連するものの遺伝子発現量解析では見逃される RNA レベルの様々な現象の動態を調べることができる。そこで本研究は、がん 1 細胞 RNA-seq データに対してリードカバレッジ解析を適用する方法論を確立することを目的とした。

【方法】 細胞型や亜型、サンプルの性質を考慮しつつ、がん 1 細胞 RNA-seq データからリードカバレッジが局所的に変動する領域 (ODEGR ; Overlooked Differentially Expressed Gene Regions) を検出する手法 multiODGERfinder を開発した。本手法ではまず各遺伝子座のリードカバレッジ行列に対して非負値行列因子分解を適用し、リードカバレッジの基本パターンを抽出した。次に、細胞集団 (細胞型・細胞亜集団など) 間で、抽出されたリードカバレッジの基本パターンのシグナルの値が変動するかを調べるため、シグナルの値に対して一元配置分散分析を適用した。さらに、一元配置分散分析で有意になった遺伝子座に対し、全ての細胞集団ペアの間でシグナルの値が変動するかを t 検定で確認することで、有意差を示す細胞集団ペアを抽出した。既報のトリプルネガティブ乳がん患者由来の 1 細胞 RNA-seq データを用いて ODEGR を人工的に導入したシミュレーションデータを構築し、multiODGERfinder の精度検証を行った。

【結果】 ODEGR を人工的に導入したシミュレーションデータに multiODGERfinder を適用することで、リードカバレッジが局所的に変動する領域を抽出できるかを検証した。その結果、multiODGERfinder は ODEGR を有する遺伝子座と ODEGR を有しない遺伝子座を判別するタスクにおいて、AUROC が 0.8 以上という高い精度を示した。さらに、複数のトランスクリプト (アイソフォーム) を含む遺伝子座を判別できるかを検証した結果、0.7 以上の AUROC を示した。このことから、multiODGERfinder によるリードカバレッジ解析が、複数の細胞集団を含むがん 1 細胞 RNA-seq データからの ODEGR の発見に寄与できることが示唆された。

1 細胞 RNA-seq データ解析の 2 つの潮流

