

論文審査の結果の要旨

氏名 河口 理紗

本論文は、RNA 2次構造の情報解析に関して、計算機による高速な2次構造予測的手法と、シーケンシングに基づく再現性のある解析手法を提案するものである。

RNA 2次構造は、細胞内でのRNAの安定性、他のRNAやタンパク質との相互作用による細胞内制御、mRNAにおけるスプライシング、翻訳に影響を与えることが知られている。RNA 2次構造を従来の実験手法で決定することは高価で効率が悪いいため、実験で決定されたエネルギーパラメータを用いた計算機による2次構造情報解析が必要とされている。しかしながら、従来の2次構造情報解析技術では、長いRNAに対して正確な計算を行うことは計算速度、メモリなどの制約で不可能であった。本論文では、長いRNAに対して整合性のある予測2次構造および様々な周辺確率の並列計算を可能とするアルゴリズムを考案し、ParasoRとして実装した。

本論文で開発されたParasoRは、Rfoldの確率モデル（確率文脈自由文法）を用い、ゲノム配列全体などの大規模な塩基配列データの2次構造解析を可能とするツールである。従来のRNA 2次構造情報解析においては、スライディングウィンドウを用いて平均化するなど精度を犠牲にした方法を用いる、もしくは短い領域を切り出して解析するしか方法がなかった。ParasoRは長いRNA中で確認された2次構造の予測に関して、従来法と同等以上の精度を示し、シーケンシング技術を用いたハイスループットな2次構造指標とも高い整合性を示した。さらに、ヒトゲノムに対する網羅的な解析によって、RNAに転写される領域は遺伝子間領域に比べて優位に2次構造を取りやすいこと、イントロン領域はその逆相補鎖や遺伝子間領域に比べて2次構造を取りやすいことなど、重要な知見が得られた。本論文で示された並列計算による厳密な確率計算は、ゲノムワイドな2次構造解析を可能とした点で、大きな意義があり、大量のゲノム、トランスクリプトームに対して網羅的で信頼性の高い2次構造情報解析が可能となる。

一方、シーケンシング技術を用いたハイスループットな2次構造指標同定技術が開発され、網羅的なRNA 2次構造解析への期待が高まっている。しかしながら、PARSやicSHAPEなどの手法間で結果にばらつきがあり、再現性に問題のあることが指摘されてきた。本論文では、信頼性と再現性に優れたRNAリアクティビティ分類を可能とする情報解析ツール、reactIDRを開発した。

本研究で示された2次構造情報解析手法は、大量のゲノム、トランスクリプトームに対して網羅的で信頼性の高い2次構造情報解析が可能であることを示すものであり、今後、幅広く活用されることが期待される。

以上1127字