

# 論文審査の結果の要旨

氏名 松本 拓高

本論文は、ゲノム配列のアセンブリにおけるハプロタイプの推定と、1分子シーケンシングによる RNA 発現データを用いた細胞分化過程の解析に関して、確率モデルを用いた手法を提案したものである。

ハプロタイプアセンブリの既存のアルゴリズムの多くは、シーケンシングされた短いリードから長いハプロタイプを決定論的に推定しているため、同一ハプロタイプと判定された領域の両端が実際に同じハプロタイプに由来する確率が必ずしも高くないという問題があった。本論文では、リードの生成過程をモデル化した混合確率モデルにおける変分ベイズ法を用いたパラメータ最適化法が提案された。この手法では、SNP の前後でのハプロタイプの連続性を確率的に評価し、既存手法より高い信頼性が得られることが示された。さらに、ハプロタイプ推定を間違える事例の解析から、ハプロタイプの異なる領域から 1 本のリードとして生成されるキメラ・リードがデータに含まれており、推定結果を低下させていることが突き止められた。この問題に対応するため、正常なリードとキメラ・リードの統計的な違いに基づくキメラ・リード判別手法を考案し、高い精度でキメラ・リードを取り除くことが出来ることが示された。また、この手法でキメラ・リード候補を除去することにより、ハプロタイプ推定の精度が向上することが示された。

従来のハプロタイプ推定においては、出来る限り長い領域を同一ハプロタイプとして予測することが多く、推定に基づく生物学的な議論に疑念を生じる恐れがあった。本論文で提示されたハプロタイプの連続性を確率によって定量的に評価する指標は、下流の生物学的解析を定量的に行うことを可能とする面で、大きな意義がある。

実験・計測技術の進歩により、遺伝子発現量 1 細胞計測、短時間間隔での計測が可能になった。本論文の細胞分化過程の解析は、このような高解像度データを活用し、遺伝子発現の詳細なダイナミクスを理解することを目指した研究である。遺伝子発現の高解像度データを解析するための計算機的手法は十分に開発されておらず、新たな手法の開発が期待されてきた。本論文では、Ornstein-Uhlenbeck 過程 (OU 過程) を用いたモデル化を提案した。ここで、一般的な OU 過程では 1 つのアトラクタしか扱えないが、OU 過程を混合化することにより、複数への分化を許すモデルを構築した。そのうえで、細胞分化の pseudo time などのパラメータを EM アルゴリズムによって推定し、分化の過程における遺伝子発現のダイナミクスを解析する新規手法を開発した。

本研究で示された細胞分化過程の解析結果は、遺伝子発現量だけから細胞系統を推定することが可能であることを示すものであり、今後、幅広い分野で活用され、発展する理論を提示したものである。

以上 1 1 6 0 字