

審査の結果の要旨

氏名 仲木 竜

本論文は、大規模エピゲノムデータを用い転写制御因子の協同性を明らかにする新規計算アルゴリズムの開発と応用に関する研究成果が、5部構成でまとめられている。第1章は、まず研究の背景として、細胞特異的な遺伝子発現制御における、転写制御因子間の協同制御機構の重要性が議論されている。有意な制御因子間の相互作用を特定する上で、あらゆる組合せを実験的に検証することは困難であり、計算科学的に確からしい相互作用を絞り込む必要がある。近年の次世代シーケンサー技術の発展に伴い、転写制御因子のゲノム DNA 認識に関わる数多くのエピゲノムデータを得られるようになり、それらを用いて転写制御因子間の協同性を予測する計算科学的手法の必要性が高まっている。本論文では、“結合領域レベル”及び“認識配列レベル”より転写制御因子の協同性を明らかにするための2つの計算手法が提案されている。

第2章では、次世代シーケンサーを用いて転写制御因子の結合領域をゲノムワイドに特定する実験手法である ChIP-seq 法より得られたデータを用いて、制御因子間の共局在性を特定する新規計算アルゴリズム (CoLo)に関する研究成果がまとめられている。近年の次世代シーケンサー技術の発展に伴い、数多くの ChIP-seq データが利用できるようになり、これまでの少数データ間の比較から、多数データを同時に比較する計算手法の必要性が高まった。少数のデータ間では、データ精度の差に基づくバイアスを個別に調整することができるが、多数データを同時する場合、それらを自動的に調整する必要がある。また、転写制御因子は、多数・少数領域、エンハンサー・プロモーター領域に結合するといった DNA 結合メカニズムが異なる。本章では、データ精度及び DNA 認識メカニズムの違いに基づくバイアスを調整し、転写制御因子間のゲノム上での共局在を特定する計算アルゴリズム (CoLo)が提案されている。CoLo アルゴリズムでは、複数データの同時比較による相対的なノイズ除去と、反復的な有意共局在の絞込みを組み合わせ、上記の多数データ間比較における2つの問題を解決している。本アルゴリズムは、基礎的な統計学的手法に対して、5倍以上多く、かつ1.4倍以上高精度に有意な共局在を特定した。転写因子 GATA2 の HUVEC 及び K562 特異的な共局在因子の特定に応用することで、既知の転写補助因子に加え、新規の補助因子が予測された。これら補助因子は、HUVEC 及び K562 特異的な GATA2 標的遺伝子の制御領域上でも共局在し、GATA2 の細胞特異的な転写活性に深く関与していることが示唆されている。更に、この

解析により特定された K562 特異的な GATA2 の共局在因子 TAL1 について、GATA2 との細胞特異的な相互作用を明らかにした。

第 3 章では、高解像度のオープンクロマチン領域データを用いることで、“転写因子フットプリント”のヘテロジェネイティーを特定する新規計算アルゴリズムに関する研究成果がまとめられている。近年、6~30bp の転写因子結合占有領域 (“転写因子フットプリント”と呼ばれる)を 1bp 単位で特定できる DNase-seq 及び ATAC-seq といった実験手法が開発された。これまで、それら実験データより転写因子フットプリントを予測する計算アルゴリズムが複数提案されたが、それらは特定の転写因子認識配列と局所的なクロマチン構造変化を一对一に結びつける解析にとどまっていた。しかしながら、異なる細胞間、または同じ細胞においても、転写因子は異なる補助因子を伴うことが知られ、その複合体形成によって転写因子フットプリントに差異が生じる可能性がある。本章では、認識配列抽出アルゴリズムと主成分分析を合わせて転写因子フットプリントのヘテロジェネイティーを特定する新規計算アルゴリズム (Hetero-DGF) が提案されている。このアルゴリズムは、これまで転写因子フットプリントの抽出において高精度を示すとされた 2 つの既存の手法と比べ、特に認識配列の短い転写因子のフットプリント抽出において 15%程高い精度を示した。また、こちらのアルゴリズムを GATA2 の HUVEC 及び K562 でのフットプリントの抽出に応用することで、細胞特異的・共通のフットプリントが特定され、それらが同じ細胞内でヘテロジェネイティーを形成していることを明らかにした。また、K562 特異的なクロマチン構造変化を伴う GATA2 フットプリントが示す認識配列パターンは、TAL1-GATA2 複合体の認識配列と一致していた。この結果は、第 2 章で特定された、K562 における GATA2 と TAL1 の細胞特異的な相互作用の結果と一致するものとなっている。

第 4 章では、大規模・多数のエピゲノムデータ解析における、本論文で提案された 2 つの計算アルゴリズムの意義が述べられている。今後とも、転写制御に関わるエピゲノムデータはより大規模となり、これら転写制御因子間の協同性を明らかにする本計算アルゴリズムの重要性が高まることが期待される。

最後に第 5 章では、以上の博士研究を進める上で参考となった先行研究に関する学術論文が一覧にされている。

よって本論文は博士 (工学) の学位請求論文として合格と認められる。