

転写制御因子の細胞特異的な協同制御機構 を解明する新規計算手法

工学系研究科・先端学際工学専攻
仲木 竜

第1章 序章

異なる機能を有す細胞の転写制御は、転写制御因子（転写因子、転写補助因子、クロマチン再構成因子等）の細胞特異的な協同制御機構により維持されている。本博士過程において、“結合領域レベル”と“認識配列レベル”の2つの観点より転写制御因子の協同制御機構を予測する、それぞれ2つの新規計算手法を開発した。

第2章 複数のChIP-seqデータの比較により共局在因子を特定する新規計算アルゴリズム (CoLo)

- 背景

近年の次世代シーケンサーの発展に伴い、直接的又は間接的にゲノム領域と相互作用する複数の転写制御因子のChIP-seqデータが得られ、多数データの同時比較が可能となった。これまで、既に複数因子が認識するゲノム領域を予測する計算アルゴリズムは提案されているが、制御因子間の細胞特異的な相互作用を予測する計算アルゴリズムは存在していなかった。

- 計算手法の特徴

異なるChIP-seqデータの比較において、実験条件の違いに基づく精度の差を調整し、各制御因子固有の結合領域分布のバイアスを取り除く事で、直接的な相互作用による制御因子間の共局在性を予測する。

- 結果

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。HUVEC（ヒト臍帯静脈内皮細胞株）及びK562（ヒト白血病細胞株）において、GATA2の既知及び新規の転写補助因子を予測した。その中で、K562特異的な共局在性を示した転写因子TAL1との相互作用に注目し、TAL1ノックダウン実験によりTAL1がGATA2の細胞特異的なDNA認識メカニズムに寄与する可能性を示した。

第3章 高解像度オープンクロマチン領域データよりヘテロジニアスな転写因子の結合フットプリントを予測する新規計算アルゴリズム (Hetero-DGF)

- 背景

DNase-seq や ATAC-seq といった高解像度オープンクロマチン領域の特定技術の発達は、1bp 単位でクロマチン構造変化を捉えることを可能とした。現在、これら高分解能エピゲノムデータを用い、転写因子の結合フットプリント (6-30bp の結合占有領域) を特定する計算アルゴリズムが複数提案されている。これらアルゴリズムは転写因子結合モチーフ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。

- 計算手法の特徴

結合モチーフ抽出アルゴリズムを用いて、局所的なクロマチン構造変化を抽出するための検索モチーフシードを特定する。その後、各検索モチーフに対応する該当領域の集合を局所的なクロマチン構造変化及び塩基配列パターンに基づき分解し、転写因子 DNA 認識における“heterogeneity”を示すフットプリント群を特定する。

- 結果

転写因子フットプリントの抽出において高精度を示す 2 つの既存の手法と比べ、特に結合モチーフ幅が狭い転写因子に対して、15%程高い精度を示した。HUVEC 及び K562 での各 DNase-seq データにおいて、転写因子 GATA2 の細胞特異的なフットプリントを特定した。これらフットプリントを分解することで、同じ細胞内でも複数種類の局所的なクロマチン構造変化を伴うフットプリントが予測され、転写因子の DNA 認識における“heterogeneity”が明らかとなった。K562 特異的クロマチン構造変化を伴う GATA2 フットプリントが示す結合モチーフパターンは、TALI-GATA2 複合体の結合モチーフと一致した。

第4章 結論

大規模・多数のエピゲノムデータ解析における上記 2 つの計算アルゴリズムの意義を考察した。今後、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、多数のデータより転写制御因子間の協同性を明らかにする上で、これらアルゴリズムの必要性が高まることが期待される。

第5章 参考文献

博士論文を作成する上で参照した先行研究に関する学術論文を一覧にした。