

# 論文審査の結果の要旨

氏名 入江 拓磨

本論文は、ヒト培養細胞中のプロモーター活性の絶対値の予測モデルの構築とその検証について述べられている。mRNAの発現量は多段階の制御を受けた産物であるため、プロモーター配列以外の影響が大きいといえる。そこで本論文では、DNA配列情報とプロモーター活性と関連付けるために、体系的なルシフェラーゼアッセイを用いた実験的アプローチをとっている。プロモーター活性予測モデルには既知の転写因子結合配列を説明変数とした線形和モデルという単純なモデルを用いているが、実測値と予測値の相関係数が $r=0.87$ をとり、比較的予測精度の高いモデルを構築することができたといえる。また、10分割交差検定の結果、相関係数の平均が $0.83$ となり未知なDNA配列へも有用なモデルを構築できたといえ、従来の転写情報のモデル化手法と比較しても精度の高いモデルであるといえる。高い予測精度を達成が可能であった理由として、定量的なプロモーター活性情報に体系的ルシフェラーゼアッセイの測定結果を利用したことが挙げられる。

in vivoな転写活性の指標としてTSS-Seqの情報を用いてプロモーター活性予測モデルのゲノム配列情報を用いた評価を行ったところ、RefSeq5'端領域のプロモーター活性予測値とTSSのタグ数との相関は見られなかったが、RefSeq5'端領域をTSSが観察された領域と観察されない領域で分類を行ったとき、プロモーター活性予測値の分布に有意な差が見られ、プロモーター領域の定性的な予測に用いることが可能であった。さらにRNAポリメラーゼIIのChIP-SeqとNucleosome-Seqの情報との比較を行った結果、TSSが観察されなかった領域で高いプロモーター活性予測値を得た領域では、RNAポリメラーゼIの結合している割合が増加し、また転写が行われている領域に特徴的な開いたクロマチン構造をとる傾向にあることがわかった。このような領域は転写産物が速やかに分解される遺伝子か、クロマチン構造を開き、Pol IIのリクルートが行われるものの転写伸長に関する因子にかけているため転写産物が見られない例であると考えられる。また遺伝子間領域においても転写産物が観察されないものの高いプロモーター活性予測値を得た領域についてはそうでない領域と比べ、Pol IIの結合、開いたクロマチン構造をとる傾向にあることがわかった。以上のことから、プロモーター活性予測モデルによって潜在的にプロモーター活性を持つ領域を予測することが可能であり、ゲノム中に潜在的にプロモーター活性を持っていると領域が多数存在していることが示唆された。ゲノム規模での転写の予測としては前段階的ではあるものの、ゲノム配列情報とトランスクリプター

ムを繋ぐ，転写制御機構の包括的な理解に向けての足がかりになると思われる。

なお，本論文は，朴 聖俊・山下 理宇・関 真秀・矢田哲士・菅野 純夫・中井 謙太・鈴木 穰との共同研究であるが，論文提出者が主体となって実験及び解析を行ったもので，論文提出者の寄与が十分であると判断する。

したがって，博士（生命科学）の学位を授与できると認める。