182　深層学習による、遺伝子発現変化原因変異の網羅的同定	王　青波

【目的】本研究では、従来手法では解釈が難しいヒトゲノムの非コード領域に存在する変異を、近傍遺伝子の細胞種特異的発現制御という文脈で包括的に理解し、その応用として人種間での遺伝子発現機構の違いや形質への影響を解き明かすことを目的とする。

【方法】ヒト 49 組織において遺伝子発現情報（RNA-seq）とゲノム情報（Whole Genome Sequence）を網羅的にカタログした公開データベースである GTEx に対して fine-mapping を適用した研究者の先行研究（Wang, Q.S. et al, Nat Commun 2021）の結果を利用し、fine-map された遺伝子発現制御変異（eQTL）を教師データとして用いた。また、特徴量として、Transformer を利用しゲノム変異が周辺エピゲノムに与える影響を予測したスコア（Enformer）等を用いた。学習器として、複数のタスク（例、人画像の性別を予測するタスクと、年齢を予測するタスク）を同時に処理することでその共通点を学び学習精度を上げるマルチタスク深層学習を用いることで、サンプル数の少ない組織や細胞種においても類似の組織から情報を学習することで高い予測精度を得ることを目指した。また、欧米での eQTL データを用い構築したスコア体系を日本人 eQTL データに適用することで、遺伝子発現制御予測スコアの適用範囲に関して評価した。

【結果】以下の工夫により、従来研究と比較し制御変異の予測精度の向上が達成されることが明らかになった。1. マルチタスク深層学習の利用：特にサンプル数の少ない脳や腎臓組織における精度向上に寄与した。2. より多くの訓練データを連続的に扱える方向へのロス関数の改良：従来の二値的な分類に加え、ラベリングの誤差を考慮した連続的なロス関数の定義により、精度の向上が達成された。3. 新規特徴量（Enformer）：主成分分析と組み合わせることで、計算コストを抑えつつ精度の向上が可能となった。同時に、新規に構築した予測器を評価可能とする、日本人集団における fine-mapped eQTL データベースを構築した（Wang, Q.S. et al, Nat Commun 2022）。該当日本人 eQTL データを利用し、予測器の予測能を評価することで、集団の遺伝的背景を超えて、機能ゲノミクスベースで高い予測能を達成可能であることが示された。

本研究の概念図