

(別紙1)

総括研究報告書

課題番号：30-37

課題名：機械学習法を用いたクロマチン高次構造の予測モデルの構築

主任研究者名（所属施設） 国立成育医療研究センター

（所属・職名）メディカルゲノムセンター/周産期病態研究部 研究員

（研究成果の要約）

1. 研究目的

近年、クロマチン立体構造の異常が小児の先天奇形や先天性疾患、悪性腫瘍の発生に関連していることが報告されつつあり、この立体構造の予測を比較的低コストで行えることが、今後の疾患関連遺伝子解析においても重要な役割を果たすと考えている。現在行われている主なクロマチン立体構造の解析は比較的高コストのため、その実施は主に研究レベルに留まっている。このため本研究に置いては、ゲノムおよびエピゲノム情報を統合し、in-silicoでのクロマチン高次構造の予測モデルを構築し、従来よりも低コストで精度の高く迅速な予測を可能とすることを主な目的とする。

2. 研究組織

研究者	所属施設
瓜生 英尚	国立成育医療研究センター
岸本 洋子	国立成育医療研究センター
青砥 早希	国立成育医療研究センター

3. 研究成果

本研究では、主に公開データベースに存在する細胞株および動物種に対して、ニューラルネットワークを基にしたアルゴリズムの構築を行った。入力ベクトルとなる情報としては、対象となる細胞のゲノム配列、マッピング情報および分化段階を評価可能なエピジェネティクスに関する情報を考慮している。クロマチン立体構造の評価項目としてループ構造の存在を指標とする。また、これらの領域の有する情報(ゲノム配列、マッピング情報およびエピジェネティクス)を揃え、これを深層学習モデルに対する訓練用のデータ(training)および評価用のデータ(test)の3群に無作為に分類する。

これらのデータセットにより、学習を終えた予測モデルの評価を行った。ループの存在の有無に関して高い感度(99.35%)、特異度(95.22%)および良好なAUC(0.997)を得られた。

4. 研究内容の倫理面への配慮

参照データベースとして使用したものは、すでに公共に公開されているもののみとし、これに関してはこのデータベース所有者がその基準に基づいて公開に至っているものであるため本研究でのデータ取扱いについては提供側の規則を遵守する以外での配慮は必要としない。