

「生命動態の理解と制御のための基盤技術の創出」
平成 28 年度採択研究代表者

H27 年度
実績報告書

武田 洋幸

東京大学大学院理学系研究科
教授

DNA3 次元クロマチン動態の理解と予測

§ 1. 研究実施体制

(1) 武田グループ

- ① 研究代表者: 武田 洋幸 (東京大学理学系研究科、教授)
- ② 研究項目
 - (1) メダカ胚からの epigenetic code、発現情報の収集
 - (2) Hi-C 法による DNA 接触位置情報取得の技術の確立と実施
 - (3) 再解読が完了した近交系 HNI ゲノムを用いて epigenetic code の系統差を同定
 - (4) epigenetic code を規定するメカニズムの解析

(2) 森下グループ

- ① 主たる共同研究者: 森下 真一 (東京大学新領域創成科学研究科、教授)
- ② 研究項目
 - (1) DNA3 次元構造における空洞を計量し、DNA と核内 body の相互作用を考察
 - (2) 上記ソフトウェアの性能を出芽酵母およびヒトの公開データを利用して検証
 - (3) DNA3 次元構造の推定を正確化するためのメダカゲノムの完備化

§ 2. 研究実施の概要

I. クロマチン構造を予測する数理モデル（主に森下グループ）

- Hi-C法で得られるDNA接触情報からDNAの3次元構造を推定し、DNA3次元構造が存在しない空洞を推定し計量するソフトウェアを、代数トポロジーの考え方（persistent homology等）を取り入れ作製した。
- 核内の空洞にどのようなタンパク質複合体（Polycomb群構造体、核スペckル）・転写ファクターが存在し、DNAと接触するかを考察するために、出芽酵母／ヒトの公開データを利用し、分析アプローチの妥当性を検証した。

II. 発生・再生過程における包括的エピゲノム情報、接触位置情報の抽出

- Hi-C法の改良版であるIn situ Hi-C法により、接触位置情報を再収集し、低ノイズで高解像度での解析を可能とするデータを得た。さらに手法の改良により初発細胞数を減らすことにも成功し、当初計画にはない初期胚由来の様々な細胞種について解析が可能となった。
- クロマチン高次構造と転写を関連付けるため、メダカ成体の whole brain、liver および fibroblast cell からのトランスクリプトームデータを stranded RNA-Seq により取得した。
- ゲノムの高次構造形成に重要な因子である CTCF の良好な抗体を得ることができ、メダカ胚期胚を用いた ChIP-Seq データを取得した。

III. Epigenetic code の確立とクロマチン構造に寄与する DNA 領域(配列)の同定（武田グループおよび森下グループ）

- DNA3次元構造を予測する際に障害となっていたメダカゲノム中に存在する約 10 万個のギャップを、483 個まで減らすことに成功した。この結果、解読配列の長さを 700Mbp から 732Mbp まで伸長させた。
- 平成 28 年度に計画している南北日本に由来する 2 種の近交系の F1 の 3 次元構造を解析するため、この 2 種と外群 1 種のゲノムを決定し、塩基変異、構造変異がどちらの種に入ったかを網羅的に明らかにした。
- メダカ近交系 2 種においてゲノム・エピゲノム比較を行い、2 種のメダカの間には塩基多型が高頻度で存在しているにも関わらず、大部分の DNA 低メチル化領域が保存されていることが明らかになった。さらに、保存された低メチル化領域において特異的に保存された DNA 配列を同定した (Uno et al., Zoological Science, *in press*)。

- これまでに、機械学習アルゴリズム (Support Vector Machine) を用いてメダカ胞胚期における DNA 低メチル化領域に特異的に存在する DNA 配列を見つけていたが、さらに詳細な解析を行った結果、これらの配列はゲノム上に周期的に存在し、ヌクレオソームのリンカー部分に一致することが明らかになった。

さらに、DNA 低メチル化領域内におけるヌクレオソーム配置は特定の DNA 配列だけから正確に予測できることも明らかにした (図6)。これまで、ヌクレオソームがどの DNA 配列を好むかはゲノム全体で解析されてきており、ゲノム領域による違いは議論されていなかった。しかし、今回の結果はクロマチンのエピジェネティックな環境によってヌクレオソームの DNA 配列への依存性が異なることを示唆している。この成果は現在論文にまとめ、投稿中である。

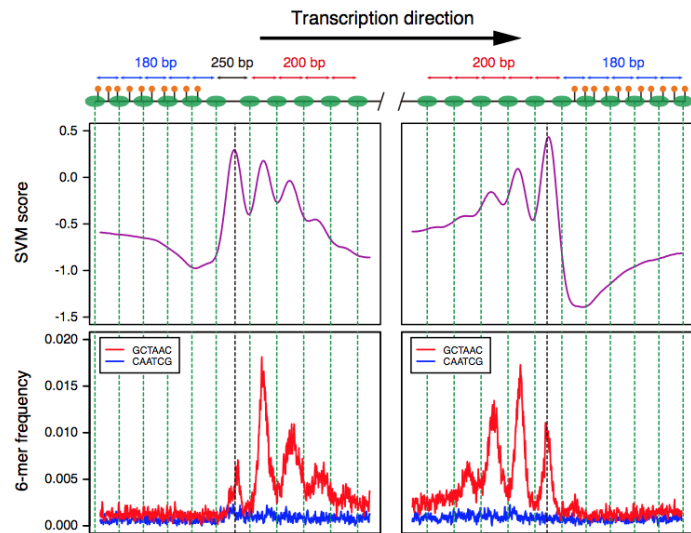


図6. ヌクレオソーム (緑楕円) のリンカー部分で SVM の予測スコアが高くなり、特定の DNA 配列 (図では GCTAAC) が高頻度で分布している。