

戦略的創造研究推進事業 CREST
研究領域「生命動態の理解と制御のための基盤技
術の創出」
研究課題「DNA3次元クロマチン動態の
理解と予測」

研究終了報告書

研究期間 2013年10月～2019年3月

研究代表者：武田 洋幸
(東京大学大学院理学系研究科、
教授)

§ 1 研究実施の概要

実施概要

本研究では、遺伝子の発現制御に重要であるクロマチン 3 次元構造が発生、分化に伴って、どのように変化し、転写能(分化能)へどのような影響を及ぼすかを明らかにする。特にゲノム解析に優れたメダカを対象に、多能性や運命決定・分化状態の維持などに重要な **developmental key genes** に注目した。また、研究後半では、初期発生の過程でクロマチン 3 次元構造が出現する過程に着目した。

研究全期間を通して、武田グループは、3 次元構造推定に必須な **epigenetic** 修飾と DNA 接触位置情報を胚、組織、細胞より網羅的に収集し、一方森下グループは、それらの情報からクロマチン構造を予測するモデルを構築した。さらに、クロマチン高次構造の解析に必須である、高品質なゲノム配列を得るために、メダカ近交系ゲノムの再解読を行った。

DNA 接触位置情報を収集する Hi-C 法は技術的に難しく、研究開始当初は欧米の限られた研究室でのみ行われていた。武田グループでは改良を加え、少数細胞で高品質な接触位置情報が得られる技術を確立した。高品質な Hi-C データが哺乳類以外の脊椎動物で得られたことになるが、その膨大な情報量の解析と意義付けに、森下グループと武田グループが共同で取り組んだ。

epigenetic 修飾の ChIP-seq による網羅的収集については、研究開始の早い段階で大きく進展し、近傍遺伝子の転写に促進的ヒストン修飾 H3K4me、抑制的ヒストン修飾 H3K27me の初期発生過程での動態が明らかとなった。その結果、**developmental key genes** は、多分化能を有する細胞で構成されている胞胚(blastula、受精後 6 時間)では、広い DNA 低メチル化領域(5 Kb から 100 Kb におよぶ広範囲)、高レベルの H3K27me、低レベルの H3K4me の集積という特徴的な **epigenetic code** で標識されていることが判明した。即ち、オープンなクロマチン状態のまま、H3K27me の集積によって強く抑制されている一方、次の原腸形成期(受精後 10-17 時間)の発現に備えて活性的修飾も有する **poised** の状態であることが判明した(Nakamura et al., *Development*, 2014)。

次に **epigenetic code** の成立機構を探るため、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、それらがヌクレオソームの配置に関与している可能性を得た(Nakamura et al., *Epigenetics Chromatin*, 2017)。一方、DNA メチル化領域に特異的配列が自律的にメチル化状態を規定しているかをメダカ胚へ導入して調べた。その結果、これら特異的配列だけは DNA メチル化は変化せず、これらの配列は他の因子とともに機能していることが示唆された(Cheung et al., *PLoS Genet.*, 2017)。

長鎖 DNA 断片(平均長 10kb)を解読できる Pacific Biosciences 社の 1 分子実時間シーケンシング技術を使い、3 つの近交系 Hd-rR, HNI, HSOK(外群)のゲノムを再解読した。DNA 配列の違いが 3 次元構造にどのような違いを生み出すかを調べる基盤を整えた。その過程で、脊椎動物モデル生物では最長のメダカのセントロメア配列を得ることができ、セントロメアの進化過程の一端を明らかにした(Ichikawa et al. *Nature Commun.*, 2017)。

DNA 3 次元折り畳み構造の基本単位である loop domain と AB compartment が、初期胚から発生過程で徐々に生成されることを観測した(後述)。このダイナミクスを理解するために loop domain (紐状の空洞) が時系列的に生成される動態を、CTCF やコヒーシン等のタンパク質複合体を考慮に入れて数理モデル化後、実装し、シミュレーションにより実データと比較し、モデルを精緻化した(投稿準備中)。

Epigenetic code とクロマチン高次構造の確立過程を明らかにするために、メダカの胚性ゲノム活性化や原腸形成過程においてクロマチン修飾および DNA 接触位置情報を網羅的に収集した。その結果、体細胞で見られるクロマチンの 3 次元構造であるコンパートメントとループドメイン(または topologically associated domain; TAD) がそれぞれ異なる発生ステージに確立されることが初めて明らかになった(Nakamura, Motai et al., 投稿中)。

(2) 顕著な成果

<優れた基礎研究としての成果>

1. Epigenetic code の成立機構に関する新しいルールが発見

概要:

機械学習アルゴリズム(Support Vector Machine)を用いて、メダカ胚期における DNA 低メチル化領域に特異的に存在する DNA 配列を見だし、さらに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なった(Nakamura et al., Epigenetics Chromatin, 2017)。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。

2. 初期発生におけるクロマチン 3 次元構造の成立プロセスの解明

概要:

大量の胚を扱えるメダカの特長と少数細胞の in situ HiC 法を組み合わせ、クロマチン 3 次元構造の成立プロセスを脊椎動物では最高の時間分解能と高解像度で解明した。その結果、体細胞で見られる階層的なクロマチンの 3 次元構造が胚性ゲノム活性化以前は存在せず、胚性ゲノム活性化以降段階的に確立されることが初めて明らかになった。

3. メダカゲノムのセントロメア構造と進化の解明

概要:

クロマチン高次構造の解析の基盤のため、モデル動物であるメダカの近交系 3 系統のゲノムを再解読した。この過程で、従来は解読が困難であったセントロメア領域の配列が、脊椎動物モデル生物では最長長さを解読できていることが判明した。解析の結果、セントロメアは染色体の端より中心部分に存在する方が、進化速度が有意に速いことが初めて明らかとなり、中心部分に存在するセントロメアが種分化に関わっている可能性が示唆された(Ichikawa et al. Nat. Commun., 2017)。

<科学技術イノベーションに大きく寄与する成果>

1. Hi-C 法の確立

概要:

Hi-C 法は非常に難しい技術とされ、本研究開始時は日本で実施している研究室はなかった。我々は新規手法である in situ Hi-C 法をメダカ培養細胞および組織試料へ適用した。特に組織試料については、これまでに数例の報告のみで非常にノイズの多い低解像度のデータであったが、ノイズの大幅な低減に成功し解像度の劇的な向上を得た。また本手法においては最低 10^6 個細胞ほどが必要とされていたが、独自の改良を加え 10^5 個細胞で行うことが可能にした。これにより、発生ステージのより早い時期や組織試料を1個体ごとに解析することが可能となった。

2. 繰り返し配列上でメチル化状態を判定するアルゴリズム AgIn 法

概要:

ゲノム上の重複配列領域(たとえば相同染色体、トランスポゾン、セントロメア等)は、1000 塩基程度の短い DNA 領域を見ただけでは重複部分間の差を識別できないという問題がある。森下グループと武田グループは米国シリコンバレーを拠点とする Pacific Biosciences 社と共同研究を進め、同社が研究開発した1分子実時間シーケンシングが出力する 10,000 塩基以上の長さの解読配列(リード)から、CpG メチル化状態を検出するアルゴリズム AgIn 法を研究開発し、2016 年に論文発表した(Suzuki Y. et al., Bioinformatics, 2016)。

3. ソフトウェア的に標的 DNA 領域を濃縮する技術

概要:

PacBio 等のロングリードシーケンシングを使った DNA 解読を進める中で、英国オックスフォードナノポア社のシーケンサー (MinION 等) を使って、DNA 中で関心のある領域だけをソフトウェア的に濃縮する技術を開発し (Masutani and Morishita, *Bioinformatics*, 2018)、手のひらサイズ PC で動作するプログラム Dyss を公開した。

<代表的な論文>

1. Kazuki Ichikawa, Shingo Tomioka, Yuta Suzuki, Ryohei Nakamura, Koichiro Doi, Jun Yoshimura, Masahiko Kumagai, Yusuke Inoue, Yui Uchida, Naoki Irie, Hiroyuki Takeda & Shinichi Morishita, Centromere evolution and CpG methylation during vertebrate speciation. *Nat Commun.*, 8,1833 (2017)

2. Napo K. M. Cheung, Ryohei Nakamura, Ayako Uno, Masahiko Kumagai, Hiroto S. Fukushima, Shinichi Morishita & Hiroyuki Takeda, Unlinking the methylome pattern from nucleotide sequence, revealed by large-scale in vivo genome engineering and methylome editing in medaka fish. *PLoS Genet.*, 13(12):e1007123 (2017)

3. Ryohei Nakamura, Ayako Uno, Masahiko Kumagai, Shinichi Morishita & Hiroyuki Takeda, Hypomethylated domain-enriched DNA motifs prepattern the accessible nucleosome organization in teleosts. *Epigenetics & Chromatin*, 10:44 (2017)

§ 2 研究実施体制

(1) 研究チームの体制について

① 「武田」グループ

- ・研究代表者: 武田 洋幸 (東京大学大学院理学系研究科 教授)
- ・研究項目
 - ・メダカ胚からの epigenetic code、発現情報の収集
 - ・Hi-C 法による DNA 接触位置情報取得 (含 in situ Hi-C 法の導入)
 - ・近交系間でエピジェネティックコードの差に寄与する配列の検証
 - ・DNaseI hypersensitive sites をゲノムワイドに収集
 - ・機械学習 SVM を用いて Epigenetic code 成立に寄与する DNA 配列抽出と検証
 - ・メダカ初期胚におけるクロマチン 3D 構造の成立過程
 - ・Hd-rR/HNI ハイブリッド F1 の成体組織 (liver と brain) の接触位置情報の取得
 - ・3D クロマチン高解像度解析のための Hd-rR/HNI ハイブリッド F1 培養細胞 (fibroblast) の接触位置情報の取得と解析

② 「森下」グループ

- ・主たる共同研究者: 森下 真一 (東京大学大学院新領域創成科学研究科 教授)
- ・研究項目
 - ・多様な細胞系譜で Hi-C を使った DNA 接触位置情報から 3 次元構造を復元し、構造の動態を解析
 - ・DNA と接触する核内タンパク質複合体・転写因子の形状/サイズ/配置を推定し、多様な細胞系譜での動態が遺伝子発現に与える影響を分析
 - ・2近交系の F1 の相同染色体間の 3 次元構造の違いを分析

- ・ 非常に長い DNA 断片 (100K-1Mbp) を部分解読しゲノム完成度をあげ、哺乳類と魚類の間の相同領域を分析し、DNA3 次元構造および TAD 構造の保存度を分析
- ・ セントロメアの配列組成と DNA メチル化パターンの関係を分析

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

- ・ Hi-C 法の開発者の一人、Erez Lieberman Aiden 博士 (Baylor College of Medicine, Houston, TX) のグループとメダカ初期胚におけるクロマチン 3D 構造の成立機構について共同研究を行った。Hi-C データは情報量が膨大で解析に時間とノウハウが必要であったが、Aiden 博士らのグループはヒト Hi-C データの解析ツールである Juicer を開発しており、このパイプラインを用いて我々のメダカ Hi-C データの解析を実施した。最終年度では論文を共同で作成した。
- ・ 理研の古関晴彦グループと Hi-C 結果の検証のためのクロマチン 3D-FISH を共同で実施した。
- ・ スタンフォード大学の Fire 教授と 2007 年以来、DNA 折り畳み構造の基本的単位であるヌクレオソーム構造について共同研究を進めている。2015 年度にはメダカゲノムの転写開始点下流において、ヌクレオソーム位置の保存度が高い領域における特異的な塩基組成と進化について報告した (Nakatani et al., BMC Genomics, 2015)。E. L. Aiden 博士との共同研究を通じて Hi-C 法がゲノムアセンブリの修正に有効であることを知った。1998 年に解読された線虫標準ゲノムには誤りが多く、Fire 教授らとロングリードでゲノムを再解読し、Hi-C 法を使ってチェックし、新しい線虫標準ゲノムを作成し、線虫研究のポータルサイト WormBase から 2018 年度末に公開予定である。また共同論文を投稿中である。
- ・ 米国シリコンバレーを拠点とする Pacific Biosciences 社とは 2012 年以来、共同研究を進めている。同社が研究開発した 1 分子実時間シーケンシングを活用した CpG メチル化検出アルゴリズム AgIn を 2016 年に発表している (Suzuki Y et al. Bioinformatics, 2016)。この AgIn 法は、相同染色体や繰り返し配列など、短い DNA 領域を見ただけでは差を識別できない重複領域における CpG メチル化の判定を可能にする新しい技術である。たとえば従来は検出が困難であった 2 倍体ゲノムの各相同染色体の DNA メチル化状態を別々に検出する方法を、オハイオ州立大学の Kin-Fai Au 教授らと確立した (Suzuki Y et al. Genes, 2018)。またセントロメア配列における CpG メチル化状態も分析でき、メダカのセントロメア配列のメチル化状態は、種分化後に安定する傾向を報告した (Ichikawa K et al. Nature Commun., 2017)。ワシントン大学の Evan Eichler 教授とは、AgIn 法を使って霊長類ゲノムにおけるトランスポソンのメチル化状態を分析する共同研究を実施した。
- ・ ドイツのマックスプランク研究所は、BLAST や Whole genome shotgun sequencing 法 (Celera アセンブラ) で知られる Eugene Myers 博士を迎え、Dresden に研究所 (Molecular Cell Biology and Genetics) を開所した。所長の Myers 博士は、PacBio の長鎖 DNA 解読装置を使った動物ゲノムの解読を幅広く進めている。セントロメアは他のゲノムでも難読領域であり、Myers 博士は我々のセントロメア解析に関心を持った。研究分担者の森下は HiC 法と PacBio シーケンシングを用いた研究アプローチを紹介し共同研究を開始した。2017 年度は博士課程大学院生 (学振 DC) を半年派遣した。
- ・ PacBio 等のロングリードシーケンシングを使った DNA 解読を進める中で、オックスフォードナノポア社のシーケンサー (MinION 等) を使って、DNA 中で関心のある領域だけを濃縮する技術に関心を持った。同社のシーケンサーをソフトウェア的に制御することで、効率的に濃縮する手法を提案し (Masutani and Morishita, Bioinformatics, 2018)、手のひらサイズ PC で動作するプログラムを公開した。