

日本医療研究開発機構 創薬等ライフサイエンス研究支援基盤事業 事後評価報告書



I 基本情報

補助事業課題名: (日本語) 創薬等ライフサイエンス研究支援基盤事業
(プログラム名) (英語) Platform Project for Supporting Drug Discovery and Life Science Research

実施期間: 平成29年4月1日～令和4年3月31日

補助事業担当者 氏名: (日本語) 白髭 克彦
(英語) Katsuhiko Shirahige

補助事業担当者 所属機関・部署・役職:
(日本語) 国立大学法人東京大学・定量生命科学研究所・所長/教授
(英語) Director/Professor, The Institute of Molecular and Cellular Biosciences, The University of Tokyo.

II 補助事業の概要

本事業では「支援」として(1)タンパク質・修飾プロファイリング、(2)ゲノム高次構造解析、(3)転写解析の3つを柱に、事業期間を通じて国内の26研究グループの研究支援を行った。(1)タンパク質・修飾プロファイリング支援ではChIP-seq・少数細胞ChIP-seq・ChIL-seq法を用いた支援を10件、サンプル数にして78の支援を行なった。(2)ゲノム高次構造解析支援ではHi-C法によるクロマチン構造の解析を5件・23サンプル行った。(3)転写解析支援では、RNA-seqによる支援を4件・36サンプル、Chromiumによる1細胞RNA-seq解析を7件・80サンプル行った。各研究支援においては、支援希望者とのコンサルティングによる実験方針の決定から試料の作製、シーケンスライブラリの調製、NGSによる配列解読、データの解析までを一貫して支援した。以下に「支援」で得られた代表的な成果を示す。

・成体ニューロン新生に対する胎生期抗てんかん薬曝露の晩発性影響とその分子基盤の解明

てんかんは脳の神経細胞(ニューロン)が過剰興奮することによってけいれんなどの発作を繰り返す神経疾患である。その罹患率は全年齢層において約1%とされており、生殖年齢の女性もその例外ではない。てんかんを合併した妊婦においては、てんかん発作の予防を目的に抗てんかん薬を継続することが原則であり、

抗てんかん薬の催奇形性に加えて、その妊娠中の投与が出生児の脳に与える長期的な影響(晩発性影響)に関する研究が盛んに行われている。晩発性影響の例として、抗てんかん薬の一つであるバルプロ酸ナトリウム(Valproic acid: VPA)の胎生期曝露による影響が挙げられるが、胎生期 VPA 曝露と出生後のけいれんの起こりやすさ(けいれん感受性)との関連は明らかとなっていなかった。九州大学大学院医学研究院の中島欽一教授らの研究グループは、脳の発生が盛んに進んでいる胎仔期に VPA の曝露を一時的に受けた成体マウスで、脳領域のうち、記憶の形成や維持に関わる海馬における新生ニューロンの移動が障害され、けいれんが起こりやすくなること、そして自発的運動によってそれらの障害が改善されることを世界に先駆けて発見した(Proc Natl Acad Sci, 2018.)。本研究における最も重要なデータの一つである、胎生期の VPA 曝露時による海馬神経幹細胞のトランスクリプトームの変化の RNA-seq による網羅的解析は、白髭グループの BINDS 支援のもと行われた。本研究の成果は、妊婦への薬剤投与が出生児の脳機能に与える影響におけるメカニズムの解明と治療法開発の一助となることが期待される。

さらに同研究グループは、脳の発生過程でニューロン産生に関わる NeuroD1 をミクログリアへ導入すると、ミクログリアの運命制御に関わるエピジェネティクスの書き換えが起こり、結果としてニューロンへのダイレクトリプログラミングが誘導されることを明らかにした(Neuron, 2019)。人為的操作により作製されたニューロンは、既存のニューロンと類似した遺伝子発現パターンを示すだけでなく、シナプスを形成することで神経回路に組み込まれ、自発的な神経活動を行う。このように、作製したニューロンは生体のニューロンと同様の性質を有していることがわかった。本研究において白髭グループは、このリプログラミングされた神経細胞の性質の評価と、NeuroD1 の具体的な作用機序を明らかにするため、ヒストン修飾と NeuroD1 の ChIP-seq 解析、RNA-seq 解析、さらに同じくバイオリジカルシーズ探索ユニットの伊藤隆司グループと連携し PBAT によるメチローム解析を BINDS の支援として行った。この成果は、損傷部位に集積したミクログリアからニューロンへダイレクトリプログラミングすることで実際に運動機能回復が見られる可能性を示しており、将来的な神経疾患治療への応用が期待できる。

・肝硬変に対するエクソソームを用いた新たな治療法の可能性

肝硬変は B 型肝炎、C 型肝炎、脂肪肝、アルコール摂取などが主な原因で長期に肝臓が障害を受け、徐々に線維化が進み黄疸、腹水、肝性脳症、静脈瘤破裂、肝細胞癌などを来す致死的な疾患で、日本に患者は 40 万人程度いる。肝臓は、元々障害を受けても再生しやすい臓器として知られているが、ひとたびこの肝硬変の状態に至った肝臓では、その線維を溶かし肝臓の機能を元にもどす非侵襲的な治療法はなく、肝臓そのものを置き換えてしまう肝移植しかないのが現状である。新潟大学・寺井崇二教授らのグループは、間葉系幹細胞から産生され非常に小さく安定な細胞外小胞、エクソソームがマクロファージを介して肝硬変に対して治療効果を発揮することを明らかにした(npj Regen. Med., 2021) 。またインターフェロン γ で刺激した間葉系幹細胞の産生するエクソソームは肝硬変に対する治癒促進効果をもつマクロファージを誘導する為に必要な物質を含むことも明らかにした。この成果は、肝硬変に対する新たな有効物質の同定や、エクソソームを用いた新たな治療法の開発につながる可能性を示している。本研究においてインターフェロン γ 処理によって得られたエクソソームの標的がマクロファージであることは白髭グループの BINDS 支援による血球の 1 細胞 RNA-seq によって明らかにされ、エクソソームによるマクロファージへの作用の詳細な解析がなされた。

・B 細胞において DNA 脱メチル化タンパク質 Tet は自己免疫疾患の抑制に働く

自己免疫疾患は、全人口の数%が罹患する難治性疾患で、これら疾患群においては全身または臓器特異的に起こる慢性炎症が長期にわたって身体機能の低下を引き起こすことで生活の質の悪化を招く。様々な自己免疫疾患が知られているが、その共通の発症メカニズムとしてリンパ球における自己寛容の破綻が考えられている。九州大学の馬場義裕教授のグループは、B 細胞に発現する DNA 脱メチル化関連酵素 Tet2、Tet3 の機能欠損が B 細胞の異常な活性化を引き起こし、脾臓等のリンパ組織に蓄積して B-T 細胞相互活性化を促す

CD86 分子を過剰に発現することを見出した。CD86 分子の過剰発現は、Tet 分子欠損によって引き起こされるエピゲノム転写抑制の破綻によって起き、この CD86 の分子機能を阻害することで、これらのリンパ球の異常活性化を抑制し自己免疫疾患病態が改善することを証明した (Nature Immunology. 2020 Aug;21(8):950-961)。本研究によって自己寛容の誘導・維持を担う基盤分子メカニズムが解き明かされたことで、そのメカニズムを標的とする新たな治療の開発が期待される。本研究における Tet の標的遺伝子の同定やそれらのエピジェネティックな制御メカニズムの解析は白髭グループによる BINDS 支援 (少数細胞 ChIP) の元に行われた。

一方「高度化」については、1. タンパク質・修飾プロファイリング、2. クロマチン高次構造解析、3. 統合解析の3つについて、我々の培ってきた解析技術の更なる高度化をおこなった。

1. タンパク質・修飾プロファイリング

通常の ChIP-seq は 10^6 ~細胞程度のインプットを用いて行われるが、さらに少ない微量の細胞のクロマチン解析のために、少数細胞 ChIP-seq プロトコルの改良と CUT&Tag 法の導入、さらにこれらに使用可能な抗体の検討を行った。その結果、最も需要の多い H3K4me3 抗体、H3K27ac 抗体で最小で 10^{-3} 細胞からでもシグナルが検出できることが確認された。さらに、H3K27me3、H3K9me3、H3K4me1、RNA polymerase, II, コヒーシン等のタンパク質においても 10^{4-6} 細胞で用いられることが確認できた。

また、さらに少量の細胞からでもデータを得るため、ChIP とは全く異なるクロマチン挿入標識法 (ChIL; chromatin integration labeling) を開発した。この方法では、細胞を化学固定後、オリゴ DNA を共有結合させた抗体で染色し、Tn5 トランスポザラーゼでその DNA 末端を抗体の標的蛋白質や翻訳後修飾の近傍に挿入する。オリゴ DNA には T7 RNA ポリメラーゼのプロモーター配列を持たせているため、オリゴ DNA の一部とゲノム配列を RNA として増幅することができる。この RNA 増幅の過程までは細胞を破壊せずに行うことができるため、単一細胞を含む極めて少数の細胞でのエピゲノム解析が可能となる。実際、ChIL-seq 法によって、ヒストン修飾の場合は 100 細胞、転写因子の場合は 1,000 細胞を出発材料としても百万細胞の ChIP-seq と同様な結果が得られた。また、H3K4me3 や H3K27ac など一部のヒストン修飾に関しては、単一細胞の解析も可能であった。また、この方法では、顕微鏡観察したサンプルの解析も可能であるため、空間的な局在とエピゲノム情報を同一サンプルで比較することも可能になった (Nat. Cell Biol. 2019.)。その後さらにこの ChIL を発展させ、同一のサンプルから2種類のタンパク質やヒストン修飾を同時に解析する multitarget ChIL 法を開発し、詳細なプロトコルを発表した (Nat. Prot., 2020)。この方法を用いることで、希少なサンプルであっても複数のエピゲノム解析を行う際にサンプルを解析ごとに分割することなく、同一のサンプルで同時に行うことが可能となった。

2. クロマチン高次構造解析

・Hi-C/HiChIP 技術の感度改善

従来の Hi-C 法及び HiChIP 法ではそれぞれ 200 万個/1 千万個の細胞が必要であるが、特定の細胞群のみに着目した解析や初期培養細胞を用いる場合などでは、これだけの細胞数を集めるのは困難である。そこで一桁から二桁の小スケール化を目指して、タグメンテーション法を用いて Hi-C・HiChIP を高感度し実験条件の検討を行った。タグメンテーション法は DNA 転移酵素トランスポザラーゼを用いることで、少量の DNA サンプルから、DNA 断片化とタグ付けをシングルステップで行い、高効率かつ短時間でライブラリ調製を行う方法である。これにより、Hi-C 法では従来の 500 分の 1 相当量の 25 ng の Hi-C DNA サンプルから解析を行うことが可能となった。また HiChIP 法においても、従来法では得られた配列データのうちおよそ 3 分の 2 が無効なリード (PCR 増幅で得られた重複リード) であったものが、タグメンテーションの導入によって約 99% をユニークなリードとすることができた。

・多点間クロマチン相互作用検出技術の開発

Hi-C 法の弱点として、DNA 断片同士の 1 対 1 の接合を基盤としているため 2 領域間の相互作用情報しか得ることができない点が挙げられる。そこで、3 箇所以上のゲノム領域の相互作用を同時に明らかにする新技術として ChIA-drop 法と Pore-C 法の開発・導入を行なった。ChIA-drop 法では、固定し断片化したクロマチンを個別に微小液滴に封入し、そこで液滴固有の配列を付加してライブラリ化することで、相互作用していた DNA 領域を再構成することが可能になる。微小液滴の作製に必要なマイクロ流体デバイスはユニット内での連携（同ユニット/竹山グループ、後述）により開発した。加えて既存のキットと液滴生成システム（10x Chromium）の利用も可能とすることで実験の簡便性・再現性の向上を果たした。また Pore-C 法では Hi-C 実験と同様の方法で調製したライブラリを断片化せず Nanopore sequencer を用いて長鎖解読を行うことで多点間相互作用をコンカテマーとして検出する。このため Pore-C 実験は Hi-C に加え特別な実験操作・実験器具への習熟を必要とせず、また Illumina 製シーケンサーに比べて低コストに解析することが可能となった。さらに Pore-C を高度化し、標的抗原周辺の多点間相互作用だけを濃縮することを可能にする、標的型 Pore-C(targeted Pore-C)の開発を進めた。これらの手法を応用してクロマチン相互作用に関わる新たな知見の探索（例：分裂時染色体構造の詳細解析、核内非膜構造複合体(phase-separation complex)の構造解析）を行った。

3. 統合解析

・ 1 細胞トランスクリプトーム解析のためのパイプライン開発

本課題で生産される 1 細胞トランスクリプトームデータを多面的・効率的に解析するため、既存の 1 細胞解析のための最新ツール群をまとめたプラットフォームを Docker イメージとして公開した (https://hub.docker.com/r/rnakato/singlecell_jupyter)。本イメージにより、遺伝子発現量、オープンクロマチン、空間的遺伝子情報などについて、品質評価、細胞クラスタリング、クラスタ特異的な遺伝子発現の同定、遺伝子ネットワーク構築などが可能である。さらなる拡張機能として、1 細胞トランスクリプトームデータを入力とした遺伝子ネットワーク推定法、細胞間で相互排他的な発現を示す遺伝子ペア群の同定法、更に複数サンプル間でのネットワーク比較を用いた新規重要マーカー遺伝子候補群の同定手法を確立し、論文化した (Nuc. Aci. Res., 2021)。本手法を用いることで、従来のクラスタ間発現変動解析では得られなかった新規のマーカー遺伝子候補を抽出することができる。

・ 複数の Hi-C データの統合解析パイプライン

Hi-C 法で得られた複数の立体構造データを比較したい場合を想定し、これまで難しかった 3 つ以上のサンプル間での相互作用変動バリエーションに基づく全ゲノムアノテーション法を開発した。本手法では比較したい複数サンプルについて、insulation score、コンパートメントスコアなど様々な 1 次元の特徴量を計算し、隠れマルコフモデルによって全ゲノム領域をクラスタリング・分類する。これによりサンプル間の立体構造摂動のバリエーションを網羅的に捉えることが可能になった。更に我々は、最近提案された様々なクロマチン状態 (stripe TAD や enhancer hub など) を捉える新規のスコアを開発した [Wang and Nakato, 論文準備中]。これらの技術を用いてコヒーシオン関連因子をノックダウンした多数の Hi-C データを比較した結果、異なるタンパク群が関与する複数の相互作用変動パターンがゲノム上に存在することが判明した。

This project has three pillars of "support": (1) protein and modification profiling, (2) genome higher-order structural analysis, and (3) transcriptional analysis. Through the project, we provided research support to 26 research groups. (1) Protein and modification profiling support included ten projects (78 samples) using ChIP-seq, low-input ChIP-seq, and ChIL-seq methods. (2) In support of higher-order genome structure analysis, 5 cases and 23 samples were analyzed for chromatin structure using the Hi-C method. (3) In support of transcriptional analysis, 4 cases and 36 samples were supported by RNA-seq, and single-cell RNA-seq analysis by Chromium was conducted for 7 cases (80 samples). For each research

support, we provided consistent support from the determination of the experimental strategy through consultation with the group who wanted support, to sample preparation, preparation of sequencing libraries, sequencing by NGS, and data analysis.

On the other hand, as "sophistication," we further pursued the advancement of the analytical techniques we have developed in the following three areas: 1) protein and modification profiling, 2) chromatin higher-order structural analysis, and 3) integrated data analysis.

1) protein and modification profiling

Although ChIP-seq is usually performed with about 10^6 ~ cells of input, for chromatin analysis of even smaller amounts of cells, we improved the low-input ChIP-seq protocol, introduced the CUT&Tag method, and investigated antibodies that can be used for these methods, and confirmed that signals could be detected from as few as 10^3 ~ cells. To obtain data from even smaller amounts of cells, we developed a chromatin integration labeling (ChIL) method, which is completely different from ChIP. This method allows comparing spatial localization and epigenomic information in the same sample in addition to low-input profiling (Nat. Cell Biol. 2019, Nat. Prot., 2020).

2) chromatin higher-order structural analysis

The conventional Hi-C and HiChIP methods typically require 2 million/10 million cells, respectively, but it is difficult to collect such a large number of cells in many cases. Therefore, we investigated the experimental conditions for Hi-C and HiChIP with high sensitivity using the Tagmentation method, aiming for a small scale of one to two orders of magnitude. As a result, the Hi-C method was able to analyze samples as small as 25 ng, equivalent to 1/500 of the conventional method. In the Hi-ChIP method, about two-thirds of the sequence data obtained by the conventional method were invalid reads (duplicate reads obtained by PCR amplification), while about 99% were unique reads by the introduction of the Tagmentation.

We also developed and introduced the ChIA-drop and Pore-C methods as new technologies to simultaneously reveal interactions between three or more genomic regions, which is not possible with the conventional Hi-C/HiChIP method. By individually encapsulating the fixed and fragmented protein-DNA complex into microdroplets and then adding each droplet-specific sequence to all the encapsulated DNA, ChIA-drop makes it possible to reconstruct the DNA regions that interacted in the genome after sequencing. In the Pore-C method, libraries prepared in the same way as in the Hi-C experiment are not fragmented, but rather long strands are decoded using a Nanopore sequencer. The multi-contacts are detected as concatemers by the sequencing. We further advanced Pore-C to develop targeted Pore-C (tPore-C), which enables enrichment of only multipoint interactions around the target antigen.

3) integrated data analysis

To analyze the single-cell transcriptome data produced by this project in a multifaceted and efficient manner, we have released a platform that brings together a set of current tools for single-cell analysis (https://hub.docker.com/r/rnakato/singlecell_jupyter). This image enables quality assessment, cell clustering, identification of cluster-specific gene expression, and construction of gene networks for expression levels, open chromatin, and spatial gene information. As a further extension, we have established and published a gene network estimation method using single-cell transcriptome data as input, a method for identifying mutually exclusive gene pair groups expressed between cells, and a method for identifying novel important marker gene candidate groups using network comparison among

multiple samples (Nuc. Aci. Res., 2021). By using this method, novel candidate marker genes that could not be obtained by conventional inter-cluster expression variation analysis can be extracted.

We have developed a whole-genome annotation method based on interaction variation among three or more samples obtained by the Hi-C method, which has been difficult before. This method calculates various one-dimensional features such as insulation score and compartment score for multiple samples to be compared, and clusters and classifies whole-genome regions using a hidden Markov model. This enables us to capture variations in 3D structural fluctuations among samples comprehensively. Furthermore, we have developed novel scores that capture various recently proposed chromatin states, such as stripe TADs and enhancer hubs. Comparison of a large number of Hi-C data with knockdown of cohesin-related factors using these techniques revealed the presence of multiple interaction variation patterns in the genome involving different protein groups.