

令和 3 年 10 月 28 日

## 海外特別研究員最終報告書

独立行政法人日本学術振興会 理事長 殿

採用年度 平成 31 年

受付番号 201960511

氏名 小林 博文

海外特別研究員としての派遣期間を終了しましたので、下記のとおり報告いたします。  
なお、下記及び別紙記載の内容については相違ありません。

## 記

1. 用務地（派遣先国名）用務地： サンフランシスコ （国名： アメリカ合衆国 ）

2. 研究課題名（和文）※研究課題名は申請時のものと違わないように記載すること。

線虫の全細胞リアルタイム in vivo 発現解析3. 派遣期間：平成・令和 31 年 4 月 1 日 ~ 令和 3 年 9 月 30 日（ 914 日間）

4. 受入機関名及び部局名

受入機関名：Chan Zuckerberg Biohub部局名：5. 所期の目的の遂行状況及び成果…書式任意 **書式任意 (A4 判相当 3 ページ以上、英語で記入も可)**

## 【記載事項】

- 研究・調査実施状況及びその成果の発表・関係学会への参加状況等
- 新型コロナウイルス感染症の影響にかかる特例措置のうち、国内採用開始・採用期間延長・翌年度渡航のいずれかの適用を受けた場合は、当該措置の適用による影響等

(注)「6. 研究発表」以降については様式 10-別紙 1~4 に記入の上、併せて提出すること。

## 【研究の背景及び目的】

多細胞生物の全細胞に対してリアルタイムに発現解析を行うことは、個体レベルでの生命現象を理解する上で重要なマイルストーンとなる。しかし、これまでの発現解析は細胞を集団としてしか扱えず、単個体さらには単細胞レベルでの発現解析は不可能であった。これまで、多数の個体の平均値としてしか得られていなかった発現データは、非常に俯瞰的な解析しかできず、個体間のばらつき、さらには細胞ごとの相互作用などまで追うことができず、より深いレベルでの生命現象の解明がはばかれていた。本研究では、単細胞レベルから単個体レベルに至るまでの発現解析を可能にする顕微法及び解析方法を開発することを目的とする。本研究は当初 *C. elegans* をモデル生物としていたが、受け入れ機関及び受け入れ研究者の都合により、ゼブラフィッシュに変更となった。しかし、変更となったものはモデル生物のみであり、その他の研究目的や研究内容はほぼ申請時のまま継承された。

本派遣では1年目の最後に、コロナウイルス感染症のパンデミックに見舞われ、光学系及びウェット系の実験を行うことが事実上できなくなった。そこで、在宅でもできる、プログラミング系の研究課題にシフトするとともに、採用期間延長が適用された。

## 【研究方法及び結果】

### 1. 光活性化 (photoconversion) 装置の開発

光活性化型蛍光タンパクは、光活性化させることで蛍光の発光波長を変えられる蛍光タンパクである。この種のタンパクは2種類あり、1つの波長で光活性化させるものと、2つの波長で光活性化させるもの (primed photoactivation) がある。2波長による光活性化は、2つの活性化ビームが空間的に重なる部分でしか光活性化が起きないため、より狭い範囲を正確に活性化させることができる。本研究で使用するライトシート顕微鏡では、励起光用の対物レンズと観察用の対物レンズが直角に配置されているため、2波長の光活性化光路を組むことができる。

本研究で開発した光活性化装置は、光学的な部分とソフトウェアの部分に分けられる。まず、光学的な部分は低コストで省スペース、且つ光照射の精度を損なわないよう、光路に特別な工夫を施している。具体的には、レーザー光源から出るレーザー光のビームスポットを小さいままガルバノスキャナに誘導することで、小さいガルバノスキャナ (低コスト、省スペース、高速) を利用できる。その後シリンドリカルレンズをうまく組み合わせることで、ビーム径を対物レンズのバックアパーチャー目一杯に広げつつ、ビームをリレーできるようにした。これにより、サンプル上に照射される光活性化のビームスポットを対物レンズ限界の小ささに保ちつつ、本来の半分のスペースで達成した。本技術は現在特許出願中である。

次に、ソフトウェアの部分では、互換性の高い Python と Qt を使って開発した。光照射する時間や強度、時空間的間隔など、光活性化に特化した機能だけでなく、このソフトウェアのウィンドウを必要に応じて半透明にし、顕微鏡のライブビューウィンドウに重ねることで、どのような顕微鏡制御システムでも利用できるようになっている。ソフトウェアからの制御自由度は高く、ユーザーがスクリーン上で描く照射パターンや模様を忠実にサンプル上で照射・再現できる (図1)。この光活性化装置を用いて実際にゼブラフィッシュの胚の一部だけを光活性化させた前後の様子を図2に示す。光活性化された細胞は最初とは異なる波長の蛍光を発するため、この部分の細胞が発生とともに、この後どのように移動していくか、正確に追跡できる。

### 2. ノイズ除去ソフトウェアの開発

*In vivo* イメージングでは、しばしば蛍光タンパクの発現量が低かったり、観察位置が深かったりなど様々な理由で蛍光の信号雑音比が低く、長時間のタイムラプスイメージングに耐えられないことがある。本研究では、*in vivo* でのリアルタイム発現解析を目標としているため、幅広い観察時間において、ノイ

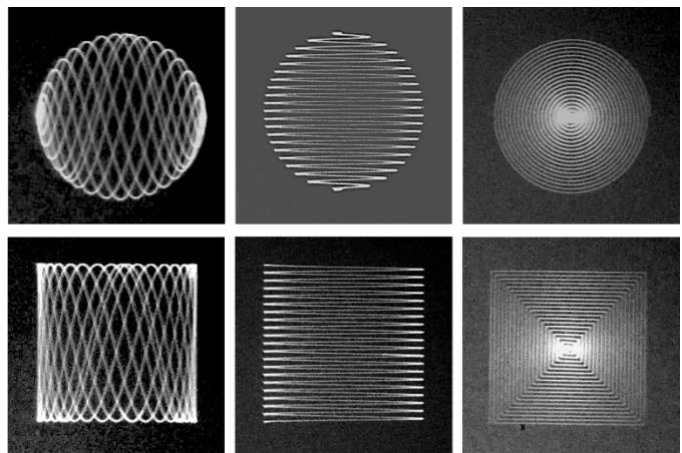


図1. 光活性化装置によって描かれる様々なパターン。光活性化装置を使って、様々な模様をサンプル上に光活性化できる。ここでは代表的なパターンを示す。見やすいよう、線の間隔を広めにとっているが、実際にはもっと密にして、選択した範囲をむらなく塗りつぶしている。

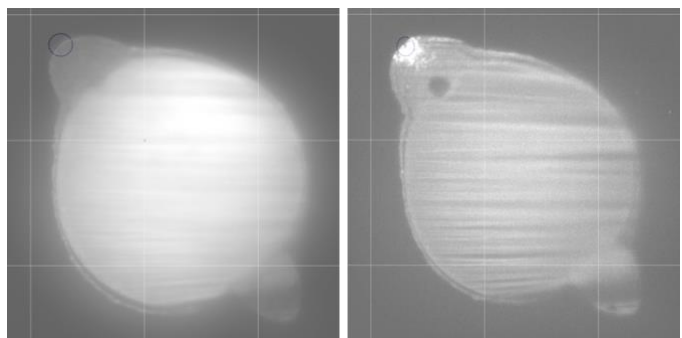


図2. 光活性化されたゼブラフィッシュ。  
(左)光活性化前、(右)光活性化後

ズの影響をできるだけ受けずに解析できることが要求される。そこで、装置開発と同時に、ノイズ除去のソフトウェア開発も同時進行で進めた。本研究が伝統的なノイズ除去と異なる最大の特徴は、機械学習を用いた **content-aware** なノイズ除去でありながら、教師なしで訓練させることである。

**Content-aware** とは、対象物の構造や位置に応じて最適なノイズ除去を施すことで、伝統的な画一的なノイズ除去と比べて機械学習の得意なタスクだが、通常ノイズの無い画像を模範として機械学習モデルをあらかじめ訓練する必要がある。しかし、ノイズのない画像を用意するのは手間がかかるだけでなく、生きたサンプルをリアルタイムで観測・解析するには適さない。そこで本研究では、教師なしでノイズ除去する機械学習モデル及びソフトウェアを開発した。本ソフトウェアはマルチ OS (Windows, MacOS, Linux) を念頭に Python で開発され、論文公開と同時にソースコードもオープンにする予定で、我々の研究のみならず、すべての人に利用・開発できるようにしてある。ノイズ除去の一例を図 3 に示す。本研究と関連する論文を arxiv に発表した。H. Kobayashi, A. C. Solak, J. Batson, L. A. Royer, "Image Deconvolution via Noise-Tolerant Self-Supervised Inversion," arXiv (2020)

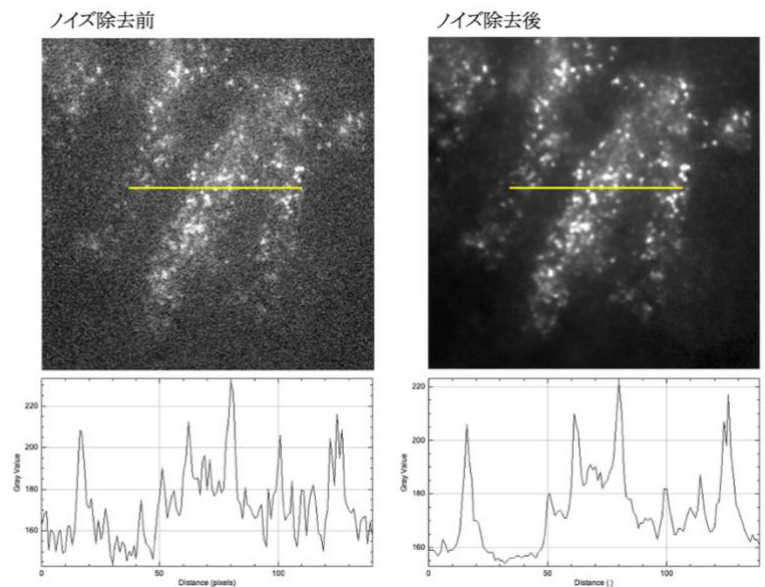


図 3. HEK293T 細胞の動画におけるノイズ除去。  
黄色い線の部分強度プロファイルを画像下に示す。

### 3. 教師なし深層学習による細胞画像のクラスタリング

深層学習は、とりわけ画像データにとって非常に強力な解析ツールであるが、大量の訓練データと正解となるラベルを必要とすることが難点であった。とりわけ、*in vivo* 画像と発現解析の双方を関連付けられる訓練データを大量に取得することは、短期間で成し遂げることはできない。教師なし学習は、正解となるラベルを使わずにコンピューターに画像の識別を学習させる手法であり、画像が撮られたい学習及び解析を開始することができる。本研究では、個別のタンパクの細胞内局在を画像で取得し、局在の画像データだけから、タンパク同士の相互作用や生化学的な類似性を画像だけから、どこまで推測できるのかを調べた。発現解析の方では様々な計算科学の手法が考案されている中で、細胞画像と生物学的現象を関連付ける研究は少なく、細胞画像から発現解析を可能にするためには、画像側の研究が欠かせない。本研究では、受け入れ機関の別の研究グループで撮影されていた細胞画像のデータを用いて始めた。

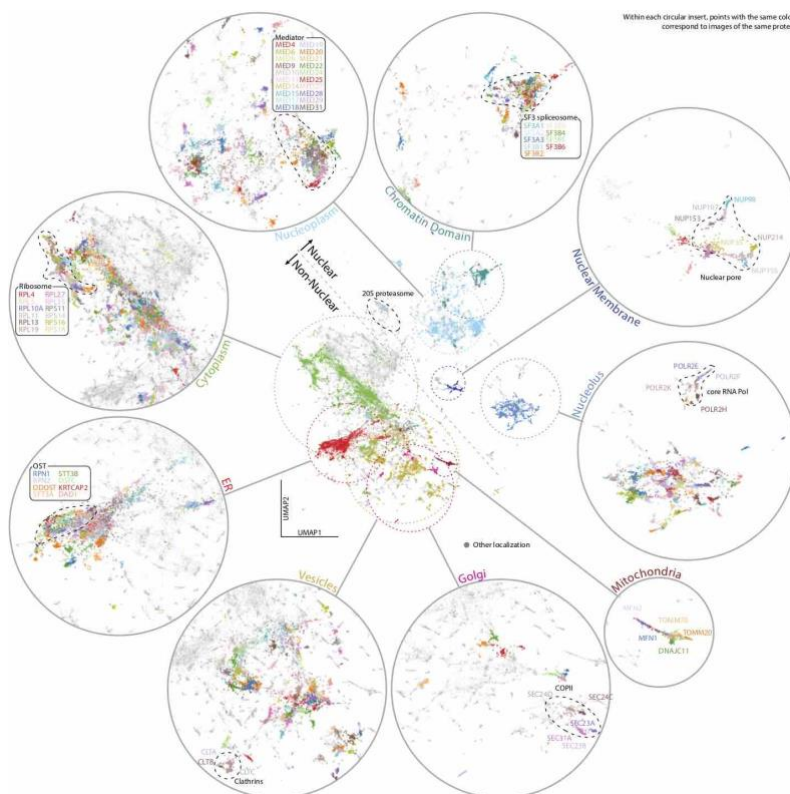


図 4. タンパク局在の高分解能アトラス。

データ点は各々の単細胞画像に対応し、細胞内での局在場所が単一のものを色分けして中央の散布図に表示し、各クラスターにズームインして、タンパクごとに色分けしたものを周りの円の中に表示してある。ここから分かるように、色を同じくする局在同士のタンパクは互いに近い場所にクラスターを作り、異なる色の局在が離れていることが見て取れる。例えば、核に局在するタンパクは軒並み右上に位置し、そうでないものは、左下に位置する。さらに、クラスターの円内をみると、タンパク複合体をつくるサブユニットがさらにクラスターを作っていることが見て取れる。例えば核小体(Nucleolus)に局在するタンパクのうち、core RNA Polymerase を構成するサブユニットが独立したクラスターを形成している。



本研究で使用した細胞画像は次のように作成された。CRISPR/Cas9 の系を用いて HEK293T 細胞の様々なタンパクを GFP で標識付けした細胞株のライブラリを構築し、spinning disk 顕微鏡を使用して細胞の 3 次元画像を取得した。1311 種類のタンパクを GFP 標識した細胞画像を約 2.4 万枚取得した。次に、画像 1 枚につき細胞が最低でも 1 つ含まれるよう、単細胞画像を切り取り、約 110 万枚の単細胞画像を準備した。本研究では、深層学習を使って細胞画像の特徴量を抽出できるように訓練し、次に抽出された特徴量を使って細胞のそれぞれのタンパクの画像的特徴によるクラスタリングを行った。深層学習の部分では、Vector Quantised-Variational AutoEncoder (VQ-VAE) をベースに深層学習モデルを構築し、抽出された特徴量を Uniform Manifold Approximation and Projection (UMAP) 二次元の散布図に可視化した (図 4)。図 4 から細胞内での局在場所が同じタンパクは UMAP 上でもクラスターをつくり、異なる局在のタンパクのクラスターと離れていることが見て取れる。さらに、各クラスターにズームインした円の中をみると、タンパク複合体を構成するタンパクサブユニットがさらなるクラスターを構成していることが分かる。これは、細胞画像だけで、どのタンパクとどのタンパクが細胞内の同じ局在場所に存在しうるかが予想できるだけでなく、潜在的に相互作用するタンパクを予測できる可能性を示唆している。画像だけで、しかも教師なし学習でタンパクの局在場所や相互作用を示唆できることを示した研究は、我々の知る限り本研究が初である。しかも類似の先行研究は核と標的タンパク以外に、細胞膜や細胞骨格をリファレンスとして染色して 4 色以上の画像を使っているが、本研究では、核と標的タンパクだけの 2 色しか利用していない。画像データだけでも、かなり豊富な生物学的情報が含まれており、これを効果的に抽出できることを本研究で示した。

本研究では、さらに一歩進んで、深層学習の解釈可能性をも推し進めた。本研究でベースにした VQ-VAE モデルは特徴量を量子化できるため、特徴量の数をカウントするアイデアを提案し、特徴量のスペクトルを作ることに成功した (図 5)。特徴量をスペクトルで表現することで、画像を表現型、特徴量スペクトルを遺伝型として捉えることができる。例えば、図 5 では、様々な細胞内局在場所に局在するタンパクの画像の特徴量スペクトルを比較すると、特徴的なスペクトルパターンをしてしている事がわかる。特に重要なことに、局在のはっきりしないタンパク (FAM241A) の画像の特徴量スペクトルをとってみると、ER のスペクトルとよく似ていることが分かる。これはこのタンパクが ER に局在していることを示唆している。このタンパクの局在を他の生化学的実験で検証した結果、確かに ER に局在していることがわかった。このことから、特徴量スペクトルが画像と特徴量の間を、遺伝型—表現型の形で橋渡ししていることがわかる。本研究とその関連する論文は bioRxiv に発表した。

H. Kobayashi, K. C. Cheveralls, M. Leonetti, & L. A. Royer, Self-Supervised Deep-Learning Encodes High-Resolution Features of Protein Subcellular Localization. bioRxiv (2021)

N. H. Cho, K. C. Cheveralls, A.-D. Brunner, K. Kim, A. C. Michaelis, P. Raghavan, H. Kobayashi, L. Savy, J. Y. Li, and H. Canaj, "OpenCell: proteome-scale endogenous tagging enables the cartography of human cellular organization," bioRxiv (2021)

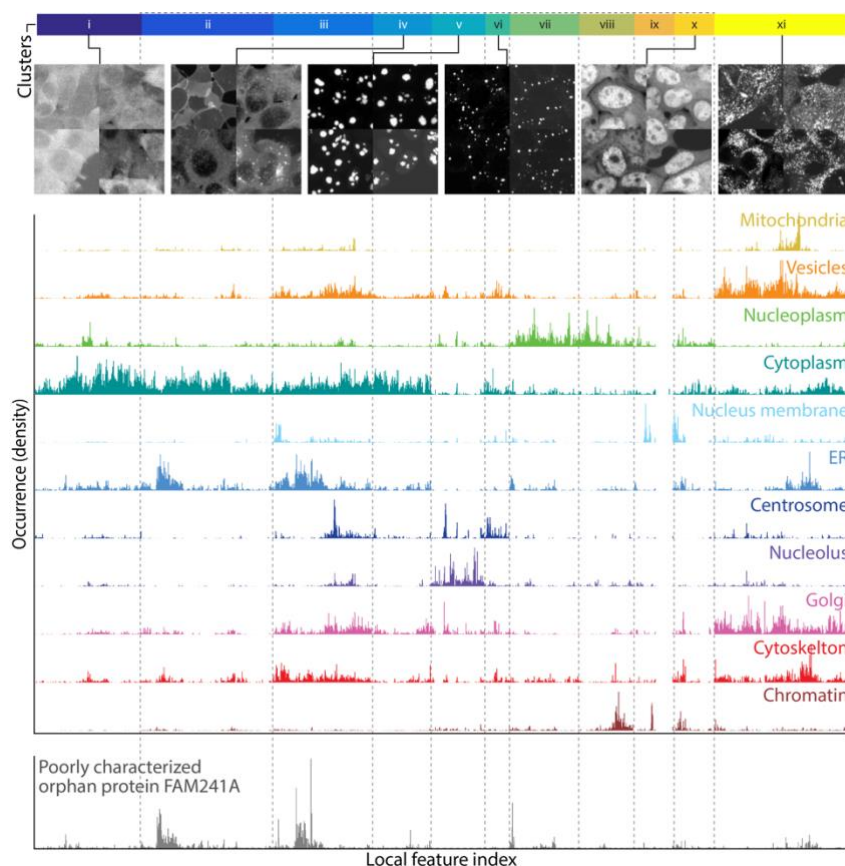


図 5. 特徴量スペクトル。

特徴量スペクトルは 2048 個の量子化された特徴量のヒストグラムである。スペクトルの縦軸は特徴量の一つの画像に存在する数を表している。特徴量スペクトルを 11 の区画に分けると、それぞれの区画が異なる画像特徴を示していることが分かる (上段)。また、細胞内局在場所ごとに分けてヒストグラムを描くと、局在場所によって特徴的な“発現領域”が見て取れる (中段)。さらに、局在の不明なタンパクの画像の特徴量スペクトルをとると、ER 局在のタンパクと非常に似ていることが分かる。このタンパクは他の実験から、確かに ER に局在していることがわかった。