

2021年 8月 5日

国立大学法人 奈良先端科学技術大学院大学  
国立研究開発法人 科学技術振興機構(JST)

## 人工知能(深層学習)により細胞内の特定タンパク質の局在推定に成功 ～関連するタンパク質のみの画像から予測可能～ タンパク質同士の相関関係の解明に期待

### 【概要】

奈良先端科学技術大学院大学(学長:塩崎一裕)先端科学技術研究科バイオサイエンス領域分子医学細胞生物学研究室の末次志郎教授(データ駆動型サイエンス創造センター/デジタルグリーンイノベーションセンター兼務)の研究グループは、細胞の画像をもとに特定のタンパク質が細胞内に局在している様子を調べる方法として、目的の2つのタンパク質間に十分に機能的な相関関係がある場合、人工知能(深層学習(注1) ディープラーニング)を用いることで、一方のタンパク質からもう一方のタンパク質の局在状況が予測できることを初めて明らかにしました。

細胞は周囲の環境に合わせて微細な構造を取り、それらの構造には特定のタンパク質の局在がみられます。これまでも細胞画像に機械学習を用いることで細胞の構造情報を抽出し、解析する試みがなされてきましたが、細胞の状態の分類や、細胞膜・核のラベリング等にとどまっており、細胞画像をもとに微細な構造の内部にあるタンパク質分子の局在を予測するといった試みはなされてきませんでした。

本研究では、深層学習の一種であるU-netモデル(注2)や、2つのネットワークにより正否を判定する敵対的生成ネットワーク Generative adversarial network(GAN)(注3)をベースにしたpix2pixモデル(注4)に、タンパク質の染色画像を学習させたところ、少なくとも機能的に関連のあるタンパク質同士では、ある一種類のタンパク質の染色画像から、他のタンパク質の染色画像を生成することができました。これを通して、コンピューターの深層学習によりタンパク質局在の推定が行えることを示しました。

1つのタンパク質の局在から他のタンパク質の局在を予測できるため、今後は細胞画像をもとに複数の他タンパク質の局在がわかるようになり、また、未知のタンパク質分子間の相関関係の解明等に役立てることができると考えています。

本研究成果は、2021年8月5日(中央ヨーロッパ時間)発行のスイス科学誌「Frontiers in Cell and Developmental Biology」のオンライン版で公開されます。

今回の研究は、奈良先端大バイオサイエンス領域の重根桂大学院生(博士後期課程2年)、スファモンジェンワタナクン大学院生(博士後期課程1年)、西村珠子助教、情報科学領域の佐藤嘉伸教授(データ駆動型サイエンス創造センター兼務)、日朝裕太助教、大竹義人准教授、スーフィーマーゼン助教らと共同で行いました。

また、研究成果はJST 戦略的創造研究推進事業 CREST 研究領域「計測技術と高度情報処理の融合に

よるインテリジェント計測・解析手法の開発と応用」(研究総括: 雨宮慶幸(高輝度光科学研究センター 理事長))における研究課題「高精度時空間計測による多元細胞情報統合」(研究代表者: 清末優子)、上原生命科学記念財団等の事業・研究領域・研究課題によって得られました。

## 【研究の背景と経緯】

細胞は周囲の環境に合わせ様々な微細構造を取ります。代表的な細胞構造には、神経突起の伸長や細胞運動等の際に形成される突起を持った構造のフィロポディア(糸状仮足)やラメリポディア(葉状仮足)(注5)が挙げられます。この構造は主にアクチン繊維という細胞骨格タンパク質によって決定されており、アクチン繊維の特定のパターンによって形成されます。すなわち、細胞構造の形成は、タンパク質の局在が鍵を握ると言い換えることができます。この特定のパターンを誘導するために、細胞構造を形成する場所には構造形成誘導タンパク質が局在しています。例えば、ラメリポディアが形成される構造の箇所には WAVE2 タンパク質や IRSp53 タンパク質が局在します。このことから、細胞構造は特定のタンパク質局在に対する情報を持っていると考えられ、近年、発展が著しい深層学習の技術等を用いることで細胞構造から情報を抽出し、特定のタンパク質の局在が予測できるのではないかと研究が進められています。しかしこれまでの研究では、細胞の構造情報やタンパク質局在情報から他のタンパク質局在を予測するという試みはなされてきませんでした。

## 【研究の内容】

本研究では、奈良先端大の末次教授らのグループにおいて、Generative adversarial network (GAN)の一種である pix2pix や U-net という深層学習のモデルを用いて、タンパク質局在推定を行いました。まず初めに、細胞構造情報であるアクチン繊維の染色画像とラメリポディアに局在する WAVE2 タンパク質の染色画像を学習させ、両者の関連を表すモデルを作成しました。このモデルを用いて、アクチン繊維の染色画像から、ラメリポディアに局在する WAVE2 タンパク質の染色画像の生成を行いました。その結果、深層学習モデルは実際の染色画像に近い画像を生成することができ、画像生成を通してタンパク質局在を予測することがわかりました(図 1)。同様の画像の生成は、WAVE2 と協調してラメリポディア等に関与する IRSp53 についても可能でした。

また、IRSp53 タンパク質の染色画像と WAVE2 タンパク質の染色画像を学習させた場合には、IRSp53 タンパク質の染色画像から WAVE2 タンパク質の染色画像を生成することができたことから、深層学習モデルによるタンパク質局在予測は細胞構造情報だけでなく関連するタンパク質の局在情報からでも可能であると考えられました。

次に、同様に、アクチン繊維の染色画像とアクチン繊維をアンカーのように結びつける小器官「接着斑」に局在するタンパク質(ビンキュリン vinculin)の染色画像を学習させたところ、アクチン繊維の染色画像から、同じように実際のビンキュリン vinculin の染色画像に近い画像が生成され、深層学習モデルはラメリポディア以外の構造からも情報を抽出し、タンパク質局在予測を行えることがわかりました(図 2)。

しかし、アクチン繊維の染色画像と細胞分裂等に関与する微小管のタンパク質(チューブリン Tubulin)の染色画像を学習させても、深層学習モデルは、アクチン繊維の染色画像からぼやけたような微小管の画像しか生成できませんでした(図 3)。

これらの結果をまとめると、深層学習モデルによる画像生成は特定の細胞構造やタンパク質のペアのみではなく、2つのタンパク質間に十分な、おそらく機能的な相関関係があれば、一方のタンパク質局在からもう一方のタンパク質局在を予測できることがわかりました。

### 【今後の展開】

深層学習を用いることで、ある一種類の染色画像から複数の染色画像が生成できるため、通常の実験では実現困難な多種類のタンパク質の染色を仮想的に行うことができます。また深層学習による画像生成は2種類の染色画像に現れた2タンパク質間の相関をもとに行っているため、この技術を用いることで未知のタンパク質の機能的な相関の解明に役立てることができる可能性が考えられます。

### 【参考図】

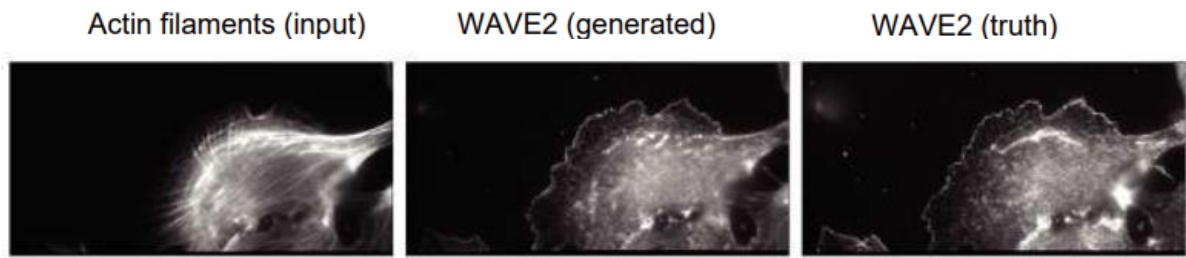


図 1: 深層学習によるアクチン繊維 (actin filament) 染色画像 (input) からの WAVE2 染色画像 (generated) 生成結果。実際の WAVE2 染色画像 (truth) も示す。

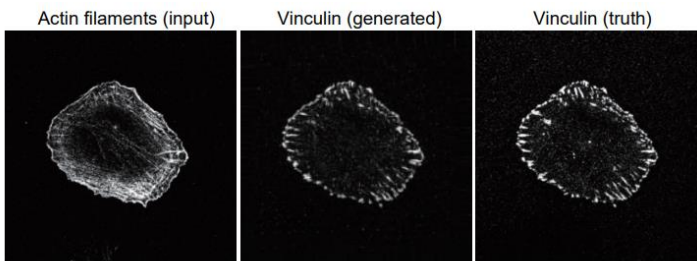


図 2: 深層学習によるアクチン繊維 (actin filament) 染色画像 (input) からのビンキュリン vinculin 染色画像 (generated) 生成結果。実際のビンキュリン vinculin 染色画像 (truth) も示す。

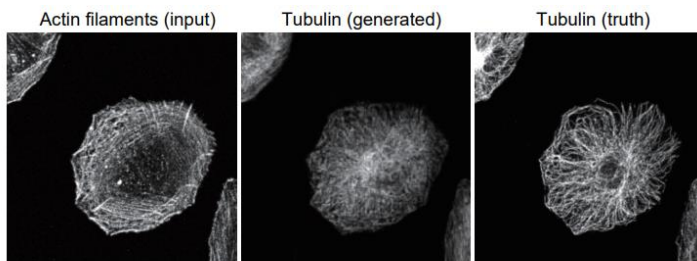


図 3: 深層学習によるアクチン繊維 (actin filament) 染色画像 (input) からの微小管チューブリン Tubulin 染色画像 (generated) 生成結果。実際の微小管染色画像 (truth) も示す。

## 【用語解説】

### 注 1) 深層学習

コンピューターに組み込まれた、生物の脳に模した人工神経ネットワークに、大量のデータ及び正解情報（画像においては対になった多数の画像）を学習させ、未知データに対して自動的にふさわしい回答を導きだせるようにする仕組みのこと。

### 注 2) U-net

U-net は 2015 年に Ronneberger らによって開発された深層学習手法である。画像のピクセルの属性を分類する方法の一つで、畳み込みネットワークにおいて、ピクセル単位で属性を分類していく際に元画像の特徴を足し合わせることで精度を大幅に向上させた手法。

### 注 3) Generative Adversarial Network(GAN)

2014 年に Goodfellow らによって提案された深層学習手法。GAN は生成器 generator と判別器 discriminator の二つのネットワークを持つ。学習には入力情報（画像等）及び目的情報（画像等）を用い、生成器は入力画像をもとに目的画像に近い画像の生成を行うよう学習し、判別器は与えられた画像が真の目的画像なのか生成器によって生成された画像かの判別を行う。このように二つのネットワークが敵対することで画像生成の精度を向上させる。

### 注 4) pix2pix

pix2pix は 2017 年に Isola らによって開発された深層学習手法である。前述の GAN をベースとしており、画像のペアをもとに学習を行うことで、一方の画像からもう一方の画像への変換を試みる。

### 注 5) フィロポディア（糸状仮足）とラメリポディア（葉状仮足）

細胞の取る微細構造の一種で、フィロポディア（糸状仮足）は、運動性が高く、細長い細胞表面の突起構造を指す。ラメリポディア（葉状仮足）は、細胞のエッジ部分に形成される扇形の構造のことを指す。両者は、細胞の運動等に関与しており、がん細胞の浸潤等の際にも形成される。

## 【論文タイトル】

“Translation of cellular protein localization using convolutional networks”

（畳み込みネットワークを用いた細胞内タンパク局在の変換予測）

DOI: 10.3389/fcell.2021.635231

## 【お問い合わせ先】

<研究に関すること>

奈良先端科学技術大学院大学 先端科学技術研究科

バイオサイエンス領域 分子医学細胞生物学研究室

教授 末次志郎

TEL :0743-72-5430

E-mail: suetsugu[at]bs.naist.jp

<報道に関すること>

奈良先端科学技術大学院大学 企画総務課 渉外企画係

TEL.:0743-72-5063 FAX:0743-72-5011

E-mail:s-kikaku[at]ad.naist.jp

科学技術振興機構 広報課

東京都千代田区四番町 5 番地 3

TEL.:03-5214-8404 FAX:03-5214-8432

E-mail:jstkoho[at]jst.go.jp

<JST事業に関すること>

嶋林ゆう子

科学技術振興機構 戦略研究推進部 グリーンイノベーショングループ

東京都千代田区五番町 7K's 五番町

TEL:03-3512-3531 FAX:03-3222-2064

E-mail:crest[at]jst.go.jp