

出芽酵母 *Saccharomyces cerevisiae* は真核生物のモデル細胞であり、約 6000 個の全遺伝子のうち約 4700 個の非必須遺伝子を個別に破壊した遺伝子破壊ライブラリーが早くから整備されたことで、それを利用した多数の網羅的な解析が行われてきた。その一つに、細胞形態と遺伝子機能との関係性を解析したのがある¹。ここでは、各遺伝子破壊株について、CalMorph と名付けられた新規プログラムを用いて、細胞形態・アクチン骨格・核の形態の蛍光顕微鏡画像データから「細胞の大きさ」や「楕円への近さ」等、501 個の形態表現型の定量的な抽出が行われた。その結果、同一あるいは類似した生物学的プロセスに属する各遺伝子機能の欠損が類似した細胞形態を生じること、すなわち、細胞の「形」はその「生理機能」を反映し、ひいては細胞の形から未解析遺伝子の機能すら予測可能であることも示された。さらに、本手法は、薬剤の標的経路の予測や酵母細胞の生理状態の把握等、様々な局面で使える強力な手法であることも示されている。このような背景のもと、私は細胞形態から、異数性のような染色体構成の異常を予測・同定することが可能なのではないかと考えた。一般的に、染色体異常はそのゲノム構成が変化してゆく不安定化状態を経て生じる。CalMorph に加え、新規な予測手法として深層学習を適用することができれば、様々な染色体異常を持つ酵母細胞のより迅速な同定や、細胞分裂に伴い染色体構成が変化している状態の細胞レベルでの追跡も可能となることが期待できる。

本研究では、様々な異数性細胞の解析を試みたが、特に解析が進んだ、二倍体 (2n) 細胞から個々の染色体について 2 本のうち 1 本を失った (2n-1) 細胞株の解析を代表例として以下に述べる。

深層学習においては、教師データによる事前のトレーニングが必須である。そこでまず、コントロールとなる野生型 2n 細胞を用いて、効率的な細胞認識手法の確立を行った。まず細胞を透過光撮影した画像から YeastSpotter² を用いて酵母細胞を抽出し、100×100 pixel の背景画像と合成したものを教師データとした。その際、出芽酵母はその細胞周期を通じて形態が変化するため、より同じ特徴を備えた画像を学習させることが望ましいと考え、細胞周期ごとに分類することとした。生産性の高い CNN モデル Xception を用いて細胞を 6 種類 [G1, S, G2/M 期, 例外 3 種 (Contamination, 複数細胞, 見切れた細胞)] に分類するよう学習を行ったところ、82% の精度で分類するモデルの作製に成功した。

2n-1 細胞株の解析にあたっては、まず、片方の染色体のセントロメア機能を人工的に阻害することで染色体の不分離に起因する欠損を誘導^{3,4} して 2n-1 候補細胞を取得し、それらの染色体構成をパルスフィールドゲル電気泳動法で確認した。本手法で 16 本ある染色体のうち、サイズの小さい染色体 (第 1, 3, 6, 9 番) については、安定な 2n-1 株が得られた。その他の染色体では、候補株は増殖遅延を示し、そこから (おそらく 2n への復帰による) 増殖速度が回復したものが一定頻度で出現した。そのため、第 1, 3, 6, 9 番染色体の 2n-1 株について、コントロール 2n 細胞と同様のモデルを用い、染色体構成の学習を行った。複数回の画像取得と学習を行ったところ、学習精度にばらつきが認められたため、写真の撮り方の誤差を認識している可能性を考えた。この可能性を排除するべく、第 1, 3 番染色体の 2n-1 株を蛍光タンパクの発現と組み合わせることで「答え合わせ」ができるように工夫し、学習させた結果を検証したところ、2n 野生型細胞と第 1, 3 番染色体の 2n-1 細胞とを混合したものについて、70% の精度で分類できていることがわかった。

現在は、CalMorph で解析に用いる画像と同じく、核・アクチン・細胞壁を蛍光染色した画像を撮影し、学習精度がどうなるかを比較解析している。

文献

1. Ohya Y et al. (2005) PNAS USA 102: 19015-19020.
2. Lu AX et al. (2019) Bioinformatics 35: 4525-4527.
3. Reid RJ et al. (2008) Genetics 180: 1799-1808.
4. Hill A & Bloom K. (1987) Mol Cell Biol 7: 2397-2405.