

【目的】 本研究では、細胞亜集団の制御手法の開発を目指し、新たな制御因子活性網羅的測定手法の開発を情動制御において重要なセロトニン神経核である背側縫線核を一例として実施した。さらに情報科学的アプローチから任意の核酸配列の制御因子活性を予測する手法の開発を実施した。

【方法】 エンハンサー活性が特に高いゲノム配列を濃縮するため、マウス背側縫線核から ATAC 法によりオープンクロマチン領域配列を単離した。単離した配列のエンハンサー活性は STARR-seq 法を用いて測定した。AAV のパッケージング後、マウス背側縫線核および前頭前皮質に投与した。4 週間の回復後、背側縫線核および前頭前皮質を単離し total RNA を回収し、逆転写反応および PCR 増幅を行ったのち次世代シーケンサーで解析した。制御因子活性の *in silico* 予測のため、まず核酸配列の前処理を行い 2,002 次元×201 ビン (200 bp/ビン) のクロマチン特徴量を算出した。このクロマチン特徴量を入力として線条体の神経細胞亜集団の遺伝子発現量を出力するような予測モデルを、WaveNet、DenseNet あるいは画像認識で一般的に用いられる 2 次元の畳み込みを多層に配置したニューラルネットワークを用いて構築した。ハイパーパラメーター探索は GPyOpt によるガウス過程最適化により行った。

【結果】 縫線核に由来する AAV ライブラリであることから、まず縫線核が含まれる中脳領域の細胞種決定に重要な転写因子群の結合配列が濃縮されているか否かを調べたところ、神経細胞の細胞種決定に重要である CTCF、Nfil3 および Rfx1 について AAV ライブラリ中で高度に濃縮されていることが明らかになった。次に背側縫線核および前頭前皮質における結果を解析したところ、両神経核でのピークは高度に類似しており、単独の神経核の解析では細胞集団選択的エンハンサー候補の単離が困難であることが示唆された。そこで、両神経核のデータを定量比較した結果、背側縫線核のみで高度に発現するエンハンサー領域を 6 か所同定することに成功した。本結果は、今回開発した手法の高い有用性を示すものと考えられた。次に、核酸配列のみからその制御因子活性を予測する手法を開発した。その結果、画像同様の 2 次元の畳み込みフィルターを多層に重ねた手法が、その他の手法よりも高い精度で予測を行えることが明らかになった。また学習に使わなかった評価データセットについて、構築した予測モデルを用いた遺伝子発現予測値と実測値の相関を調べたところ、そのスピアマンの相関係数は 0.607 ($P=2.26 \times 10^{-9}$) と有意な相関を示すことが明らかになった。これらの結果は、本予測手法の高い性能と妥当性を示すものと考えられた。今後、前項で得られた制御因子活性の網羅的測定手法と組み合わせるさらなる学習を行わせることで、あらゆる核酸配列の制御因子活性の予測が可能になると考えられる。

網羅的エンハンサー活性測定法による亜集団選択的エンハンサー配列の濃縮

