

BERT を用いた T 細胞受容体の機能解明

安水 良明

大阪大学 大学院医学系研究科

1. はじめに

CD4⁺T 細胞は免疫の中心的役割を果たしており、その異常は自己免疫疾患・腫瘍など、様々な疾患の原因となる。申請者等は自己免疫疾患における CD4⁺T 細胞の関わりを調べるため、シングルセル解析と呼ばれる 1 細胞から遺伝子発現や T 細胞受容体(TCR)配列を検出実験手法を用いた解析をおこない、Treg らしさや細胞傷害性など、CD4⁺T 細胞を形作る 12 個の特徴量を定義することに成功している。

抗原認識を行う TCR は DNA 再構成により 2×10^{19} 種類にも及ぶ多様性を有しており、疾患における意義が重要視されてきた一方で、T 細胞受容体配列の持つ意義はほぼ明らかになっていなかった。そこで本研究では、シングルセルデータを元に、TCR 配列の特徴のみから CD4⁺T 細胞を形作る 12 個の特徴量の予測を行い、どの CD4⁺T 細胞特徴が TCR によって形成されるかを明らかにすることを目的とした (図 1)。

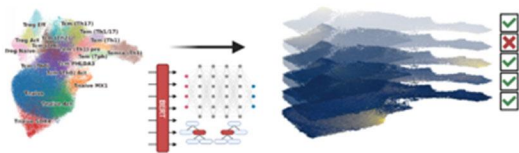


図 1 : TCR を用いた CD4⁺T 細胞特徴量の予測

2. 手法

本研究では、CD4⁺T 細胞の持つ TCR 配列について BERT の事前学習をおこない、CD4⁺T 細胞で事前学習を行った BERT モデルで分散表現抽出を行った後、全結合層や勾配ブースティングなどによる学習モデルの構築を行った。同モデルを用いて、T 細胞特徴の予測を行った。

3. 結果

はじめに、Bulk TCRseq による 500 万 TCR 配列をもちいて Treg らしさの予測を行ったところ、線形モデルを用いた既報に比べ良好な予測に成

功した。次に、CD4⁺T 細胞特徴の予測を行うため、遺伝子発現量と TCR 配列がわかっている 86,634 細胞で学習器 (TCR-BERT + lightGBM) を作成、新たに用意した 10,412 細胞を用いて評価を行った。その結果、memory らしさ、Naive らしさ、細胞障害性が TCR により説明できることが明らかになった (図 2)。

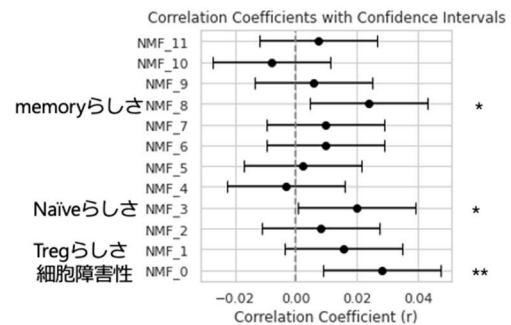


図 2 : 予測結果

4. おわりに

本研究では TCR-BERT をもちいた TCR 特徴解析は既存の線形手法に比べて有用であることを明らかにした。更に T 細胞極性 (Treg, Tfh, Th1 など) にくらべて、メモリーらしさ、細胞障害性、ナイーブらしさなど、大区分的な形質のほう TCR による影響を受けていることを明らかにした。

本研究では TCR 特徴の持つポテンシャルを明らかにすることが出来た。今後は患者検体を用いた大規模データ構築、実験的検証などを組み合わせてよりヒト疾患に即した研究を行うことで、社会実装につなげることが期待される。

参考文献

- (1) Yoshiaki Yasumizu *et al.*, “Single-cell transcriptome landscape of circulating CD4⁺ T cell populations in autoimmune diseases” *Cell Genomics*, 2024, 4, 2, 100473, 2024.