

**【目的】** がんはゲノムの変異を要因とする病気であり、がんの予防、診断、治療方針の決定のためには、個々人のゲノム配列を理解することが重要である。本研究では、現在理解の進んでいない非コード領域のゲノム配列変異とがんの関係について着目し、深層学習を用いた非コードゲノム配列の解析により、がん化における新規ドライバー変異を探索するアルゴリズムの開発を行うことを目的とした。

**【方法】** 本研究では、独自開発したゲノム配列を読み込む専用の畳み込みニューラルネットワーク (FRSS) を応用して、癌ゲノムの非コード領域の体細胞変異が、エピジェネティックシグナルに及ぼす影響を予測する方法を検討した。癌ゲノムとしては、TCGA の 1,072 の乳がんゲノムについて、small nucleotide variants (SNVs) の情報を利用した。まず、テスト段階として、畳み込みニューラルネットワークモデルをマウスの複数の正常組織、細胞株における CCCTC 結合因子 (CTCF) の ChIP-seq シグナルを用いて訓練した。癌細胞における SNVs の情報をゲノム配列に変換し、訓練したモデルを用いて、体細胞変異が入った癌ゲノムとコントロールとして体細胞変異が入っていないレファレンスゲノムに対して、CTCF 結合領域の予測を行い、予測結果を比較した。

**【結果】** まず、テクニカルな側面として、レファレンスゲノムと癌ゲノムにおける結合予測を比較する上で、小さい DNA 欠損や挿入に起因するゲノム座標のズレが、比較を困難にすることが分かった。この問題に対応するために、現在、全ゲノム配列アラインメントによってゲノム座標の補正を行っているが、より精度の高い座標補正方法が今後の課題であることがわかった。しかしながら、現在の手法から、体細胞変異に起因した、650 箇所の CTCF 結合領域の獲得と 682 箇所の欠失を予測することが出来た (下記の図、赤い部分)。今後は、CTCF のみならず、エピジェネティック因子の変化を網羅的に予測し、実験的な検証およびがん化における意義について、検討して行く予定である。

がんゲノムにおける CTCF 結合領域の獲得と欠損の予測結果

