E-news
R3-No. 9 (31/Aug/2021)

研究紹介

AI で新型コロナウイルス感染症に関連する遺伝子群を特定
～バイオインフォマティクスでコロナに打ち勝つ基礎づくり～

工学科エネルギー環境工学コース・助教　宮田　龍太
（E-mail：miyata26@tec.u-ryukyu.ac.jp）

１．はじめに

現在、新型コロナウイルス感染症（COVID-19）が世界中で猛威を奮っています。SARS-CoV-2 により引き起こされるこの感染症は 2019 年 12 月に中国で最初に確認され、瞬く間に多くの人命を奪い、私たちの日常を大きく変化させました。現状を打開すべく、様々な研究機関でワクチン開発や診断・治療法の確立をめざして基礎研究、後遺症の調査など数多くの治験や臨床研究が行われており、COVID-19 の詳細なメカニズムを解明することが急務となっております。

COVID-19 特効薬開発のヒントを得る方法の一つとして、遺伝子発現解析があります。この解析は患者さんと非患者さんで遺伝子の発現レベルがちがう箇所を調査し、疾患発症に関わる遺伝子をできるだけ少ない個数で特定することが目的です。遺伝子発現解析の難しさは、候補となる遺伝子の数が数万個と膨大になる一方、サンプル（データ）数は採取するコストが高いため、数個しか手に入らない large p small n 問題にあります。今回使用したデータも候補となる遺伝子数は 59,618 個でサンプル数は 34 名分でした。

２．方法と結果

そこで筆頭著者の藤澤氏らは、「主成分分析に基づいた教師なし学習による変数選択法（PCA-based unsupervised feature extraction, PCAUFE1）」という AI を使い今回のゲノムデータが抱える large p small n 問題を解決しました。この AI 手法を使うと、全くウイルスに感染していない通常の人と比べて新型コロナウイルス感染者の体内で発現量が異常に増えている/減っている遺伝子を検出できます。被験者の血液から採取した遺伝子発現量データ 2 に PCAUFE を適用し、123 個が COVID-19 発症に関連する遺伝子群として特定できました。

これらの遺伝子が持つ機能をバイオインフォマティクスのデータベース 3 を使って調べたところ、免疫に関連するものが多く含まれていることがわかりました。さらに、PCAUFE で特定した 123 個の遺伝子の発現を制御する転写因子をバイオインフォマティクスのデータベース 4 を使って検索したところ、NFKB1 と RELA が上位にヒットしました。これらは「防御システムの要」と言われている NF-κB に深く関係する転写因子であり、それらの活性がヒストン修飾 H3K36me3 で抑制されていることがわかりました（図 1）。

３．おわりに

一連の解析により新型コロナウイルス感染症でヒトの免疫機能が低下するメカニズムの一端を解明しましたが、はたして今回特定した 123 個の遺伝子群に単なる相関関係に留まらず COVID-19 発症の「因果」と呼べるものが含まれているかどうか更なる検証を重ねる必要があります。課題は山積みですが、効果的な治療法の確立には臨床試験のみならず、バイオインフォマティクスをはじめとした多様な病態解析で知見を積み重ねることが重要だと著者一同は考えております。

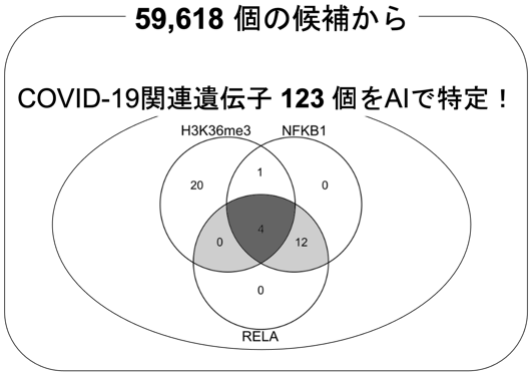

図 1：研究成果の概要

＜参考文献＞
1. Taguchi, Y-h. (2020) Unsupervised feature extraction applied to bioinformatics: a PCA based and TD based approach. Springer.
2. Arunachalam, P. S., et al. (2020) Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science 369: 1210–1220.
3. Zhou, Y., et al. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. 10: 1523.
4. Chen, E., et al. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14: 128.

＜謝辞＞
・本研究は、沖縄県委託事業である令和 2 および 3 年度「健康・医療産業における情報技術活用促進事業」の一環として実施されました。
・今回使用したデータは全て National Center for Biotechnology Information (NCBI) の Gene Exression Omnibuns (GEO) から取得しました。

＜論文情報＞
1. 論文タイトル　PCA-based unsupervised feature extraction for gene expression analysis of COVID-19 patients（COVID-19 患者の遺伝子発現量解析への PCA に基づいた教師なし学習による変数選択法の適用）
2. 雑誌名　Scientific Reports
3. 著者　Kota Fujisawa, Mamoru Shimo, Y-h. Taguchi, Shinya Ikematsu, and Ryota Miyata*
（藤澤 孝太[1], 志茂 衛[2], 田口 善弘[3], 池松 真也[4], 宮田 龍太[5]
[1] 東京工業大学生命理工学院, [2] 琉球大学大学院理工学研究科, [3] 中央大学理工学部,[4] 沖縄工業高等専門学校生物資源工学科, [5] 琉球大学工学部）
4. アブストラクト URL　https://www.nature.com/articles/s41598-021-95698-w