



令和3年8月27日  
琉球大学  
中央大学  
沖縄工業高等専門学校

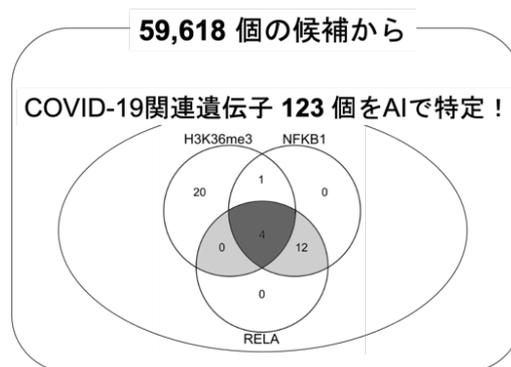
## AIで新型コロナウイルス感染症に関連する遺伝子群を特定

～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～

琉球大学工学部宮田龍太助教、中央大学理工学部田口善弘教授、沖縄工業高等専門学校生物資源工学科池松真也教授らのチームで取り組んだ、藤澤孝太氏の琉球大学大学院理工学研究科博士前期課程の修士論文テーマである本研究成果がNature Research社の学術雑誌「Scientific Reports」誌に掲載されました。内容等については別紙の通りになりますので、よろしくお願い申し上げます。

### <発表のポイント>

- ◆ **どのような成果を出したのか**：特効薬開発の取っかかりを得るために、新型コロナウイルス感染症に関連する遺伝子群123個を約6万個の候補の中からAIで特定。さらに、これらの遺伝子の発現を上流で制御する転写因子をバイオインフォマティクス<sup>(注1)</sup>の知識を活用して特定。
- ◆ **新規性(何が新しいのか)**：今回特定した転写因子<sup>(注2)</sup>には免疫防御システムの要と言われるものが含まれており、その活性が抑制されていることが示唆され、新型コロナウイルス感染症によりヒトの免疫系機能が低下するメカニズムの一端を解明。
- ◆ **社会的意義/将来の展望**：世界中で猛威を奮う新型コロナウイルス感染症の全容を解明するには、臨床試験だけでなく本論文のようなゲノムデータ解析をはじめとした多様な病態解析で知見を積み重ねていく必要があります。今後は重症化や変異株に関連する遺伝子を探索していく予定です。



### 【問合せ先】

琉球大学工学部工学科エネルギー環境工学コース  
助教 宮田 龍太  
TEL：098-895-8620  
E-mail：miyata26@tec.u-ryukyu.ac.jp

<発表概要>

● 誰がどのような成果を出したのか

琉球大学工学部宮田龍太助教、中央大学理工学部田口善弘教授、沖縄工業高専門学校生物資源工学科池松真也教授らの共同チームが、COVID-19に関連する遺伝子群 123 個を約 6 万個の候補の中から AI で特定した。さらに、これら 123 個の遺伝子の発現を上流で制御する転写因子をバイオインフォマティクス<sup>(注1)</sup>の知識を活用し特定した。今年の 3 月に大学院理工学研究科博士前期課程を修了した藤澤孝太氏の修士論文テーマであった本研究成果は、Nature Research 社から刊行された国際誌「Scientific Reports」のオンライン版に掲載される予定である。

● 新規性 (何が新しいのか)

特定した転写因子<sup>(注2)</sup>には防御システムの要である NF- $\kappa$ B(図1の NFKB1 と RELA)が含まれており、さらにそれらの活性がヒストン修飾<sup>(注3)</sup>で抑制されていることが解析結果から示唆され、新型コロナウイルス感染症によりヒトの免疫系機能が低下するメカニズムの一端を明らかにした。

● 社会的意義/将来の展望

いま世界中で猛威を奮っている新型コロナウイルス感染症の全容を解明し、効果的な治療法を確立するには、臨床試験だけでなく本論文のようなゲノムデータ解析をはじめとした多様な病態解析で知見を積み重ねる必要があり、今後チームは重症化や変異株に関連する遺伝子を探索していく予定である。

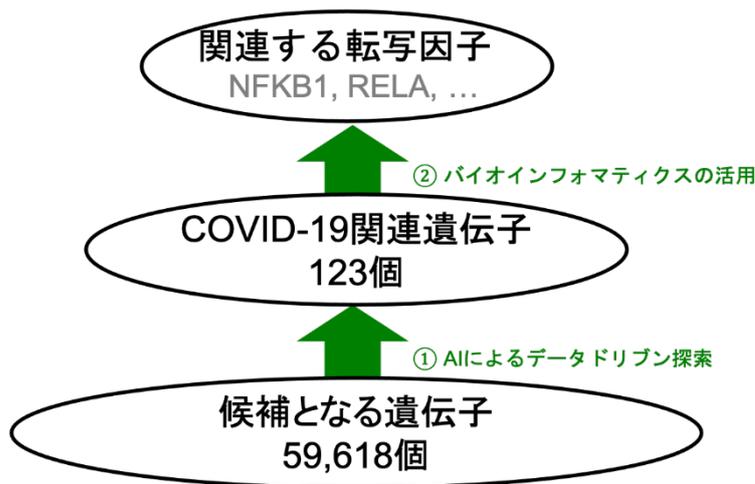


図1: 本論文の概要 (論文本編 Fig. 1 を改変)。



## PRESS RELEASE

<発表内容>

### ① 研究の背景・先行研究における問題点

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・治療法の確立にむけた基礎研究、後遺症について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 特効薬開発の手がかり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な一方で、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という large p small n 問題<sup>(注4)</sup>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 研究内容（具体的な手法など詳細）

筆頭著者の藤澤氏は、中央大学の田口教授が開発した「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, PCAUFE<sup>1</sup>）」を使って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 手法を適用することで、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。PCAUFE を被験者の血液から採取した遺伝子発現量データ<sup>2</sup>に適用したところ、表 1 に示す 123 個が COVID-19 の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 Metascape<sup>3</sup> というバイオインフォマティクスのデータベース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 Enrichr<sup>4</sup> というバイオインフォマティクスのデータベースを使って検索したところ、図 2 の通り、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- $\kappa$ B 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。

## PRESS RELEASE

### ③ 社会的意義・今後の予定

一連の解析で、新型コロナウイルス感染症によりヒトの免疫系機能が低下するメカニズムの一端を解明できましたが、はたして今回特定できた 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、COVID-19 の効果的な治療法を確立する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ていくことが重要だと著者一同は考えております。

上記と併せて、今後は COVID-19 重症化や変異株に関連する遺伝子群を探索する予定です。

表 1：本研究で特定した COVID-19 関連遺伝子群 123 個。

ACTB	ACTG1	ADRBK1	AHNAK	ALAS2	ANXA1	ANXA2	APLP2	ARL4C	B2M
BTG1	BTG2	C1orf63	CCR7	CD14	CD163	CD69	CD74	CD83	CLU
COX1	CTSB	CTSS	CXCR4	CYFIP2	DDX3X	DDX5	DNAJB1	DUSP1	DUSP2
EEF1A1	EIF1	EIF4G2	ENO1	F13A1	FCN1	FLNA	FOS	FOSB	FTL
GAPDH	GLUL	GPR183	GRN	HBA1	HBA2	HBB	HLA-B	HLA-DPA1	HLA-DRA
HLA-DRB1	HLA-DRB5	HLA-E	HMHA1	HSP90B1	HSPA5	HSPA8	IFI27	IFITM3	IGJ
IL10RA	IL1B	IRF1	ISG15	ITGA2B	ITGB2	IVNS1ABP	JAK1	JUNB	JUND
KLF2	KLF6	LCK	LCP1	LOC100507709	LOC100507714	MAFB	MCL1	MX1	NFKBIA
NFKBIZ	NR4A1	NR4A2	PIK3IP1	PKM2	PLBD1	PNRC1	PPBP	PPP1R15A	PSAP
PTGER4	PTPRC	RGS2	RPL13	RPL3	RPS2	S100A12	S100A8	S100A9	SELL
SERPINA1	SF3B1	SH3BGRL3	SLC2A3	SORL1	SPARC	SRSF5	SRSF7	SUN2	TAGAP
TLN1	TMEM66	TNFAIP3	TNFRSF1B	TSC22D3	TUBA1A	TYMP	UBC	VCAN	YPEL5
ZFP36	ZFP36L2	ZNF331							

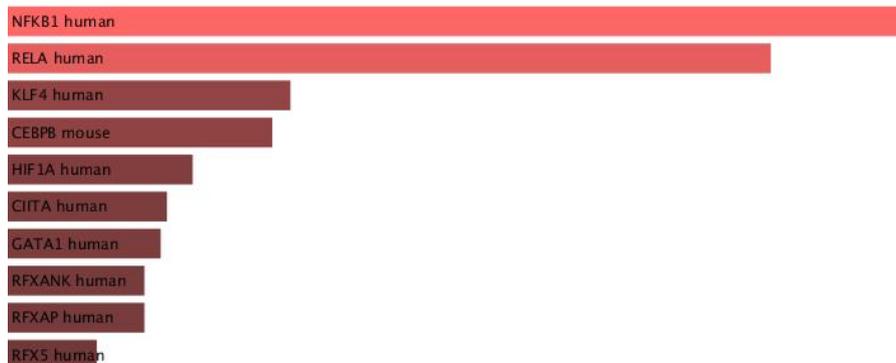


図 2：表 1 の COVID-19 関連遺伝子 123 個の発現を制御する転写因子の候補群 (Enrichr TRRUST Transcription Factors 2019 より)。棒グラフの長さや色の明るさはフィッシャーの正確確率 (検定の P 値の小ささ) を表現している。



## PRESS RELEASE

### <用語解説>

- (注1) バイオインフォマティクス：生命科学と情報科学の融合分野のひとつで、DNA や RNA、タンパク質をはじめとする生命がもつ情報をコンピュータで分析することで、生命現象を解き明かすことを目的としている。
- (注2) 転写因子：DNA に特異的に結合するタンパク質で、ある特定の遺伝子の転写（DNA の情報を RNA に写しとる過程）レベルを増減させる機能をもつ。
- (注3) ヒストン修飾：細胞の核に存在するタンパク質であるヒストンはアセチル化やメチル化など様々な化学的な修飾を受けており、それが遺伝子発現制御に関与していると考えられている。
- (注4) Large p small n 問題：サンプルの数が少なく、変数の数が多いデータを扱う場合に生じる問題で、通常の統計手法による解決は困難。数学の連立方程式で、式の数よりも未定係数の数が多いと解が1つに定まらない状況によく似ている。

### <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 Expression Omnibuns (GEO) から取得しました。



## PRESS RELEASE

### <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<sup>1</sup>, 志茂 衛<sup>2</sup>, 田口 善弘<sup>3</sup>, 池松 真也<sup>4</sup>, 宮田 龍太<sup>5</sup>  
<sup>1</sup>東京工業大学生命理工学院, <sup>2</sup>琉球大学大学院理工学研究科, <sup>3</sup>中央大学理工学部, <sup>4</sup>沖縄工業高等専門学校生物資源工学科, <sup>5</sup>琉球大学工学部)
- (4) DOI 番号 10.1038/s41598-021-95698-w