

テンソル分解に基づく教師なし特徴抽出による 新型コロナウイルス (SARS-CoV-2) に対する 新しい先進的 *in silico* 創薬手法の提案

田口 善弘^{1,a)} ターキー ターキー²

概要 :

背景:COVID-19 は、世界中の人間社会に影響を及ぼしている重要なパンデミックであり、有効な薬剤の開発が急務となっている。COVID-19 の治療に有用と思われる薬剤候補化合物は多数存在するが、その評価には時間とコストがかかる。したがって、実験的な検証に先立ち、有効な可能性のある薬剤を同定するためのスクリーニングが必要である。

方法:本研究では、重症急性呼吸器症候群コロナウイルス 2 に感染した複数の肺がん細胞株の遺伝子発現プロファイルに、最近提案されたテンソル分解 (TD) ベースの教師なし特徴抽出 (FE) を適用した。TD-based unsupervised FE により選択された 163 遺伝子の発現を有意に変化させる薬剤候補化合物を同定した。

結果:多数の薬剤のスクリーニングに成功し、その中には、C646, chelerythrine chloride, canertinib, BX-795, sorafenib, QL-X-138, radicicol, A-443654 などの多くの抗ウイルス剤既知化合物が含まれていた。CGP-60474, アルボシディブ, ミトキサントロン, QL-XII-47, ゲルダナマシン, フルチカゾン, アトルバスタチン, ケルセチン, モテキサフィンガドリニウム, トロバフロキサシン, ドキシサイクリン, メロキシカム, ゲンタマイシン, ジプロモクロルメタンがありました。また、抗寄生虫薬として同定され、最近では SARS-CoV-2 の臨床試験に含まれるようになったイベルメクチンもスクリーニングで同定された。

結論:我々の戦略でスクリーニングされた薬剤は、COVID-19 患者の治療に有効な候補となる可能性がある。

1. はじめに

コロナ禍は一時の凄惨な状況を脱し、mRNA ワクチンを始めとしたワクチンの開発と普及によって、社会活動が全面的に停止するような状況ではなくなっている。しかし、根本的な治療薬がないために、その影響から世界は完全に脱することが出来ない。

創薬は費用と時間のかかる困難なプロセスであり、コロナ禍の非常に初期から、このプロセスをコンピュータの力をつかってわずかでも短縮しようといういわゆるインシリコ創薬の試みは行われてきた [2]。

インシリコ創薬は、リガンドに注目する Ligand Based Drug Design (LBDD)、タンパクに注目する Structure Based Drug Design (SBDD)、そして薬物を投与した時の生命体の反応を基準にスクリーニングを行うフェノタイプ

ベースのスクリーニングなどがある。

この研究報告では、テンソル分解を用いた教師なし学習による変数選択法 [3] を COVID-19 を引き起こすウイルスである SARS-CoV-2 を感染させたヒト肺培養細胞の遺伝子発現プロファイルに適用し、感染に関係している遺伝子を選び、その遺伝子の発現に投与が影響を与える低分子化合物をデータベースから検索することでドラッグリポジショニングを行った研究 [1] を報告する。

2. 方法

図 1 は全体のフローチャートを示す。

2.1 遺伝子発現プロファイル

遺伝子発現プロファイルは、GEO から GSE147507 [6] の ID でダウンロードした。用いたのは GSE147507.RawReadCounts.Human.tsv.gz というファイルで、これは 5 種類の培養細胞 (Calu3, NHBE, A549 Multiplicity of infection (MOI) 0.2, A549 MOI 2.0, and A549 ACE2 expressed) に SARS-CoV-2 の感染と非感

¹ 中央大学
Chuo University, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

² キング・アブドゥルアズィーズ大学

^{a)} tag@granular.com

本研究は原著論文として刊行済みである [1]

Evaluation, Drug repositioning

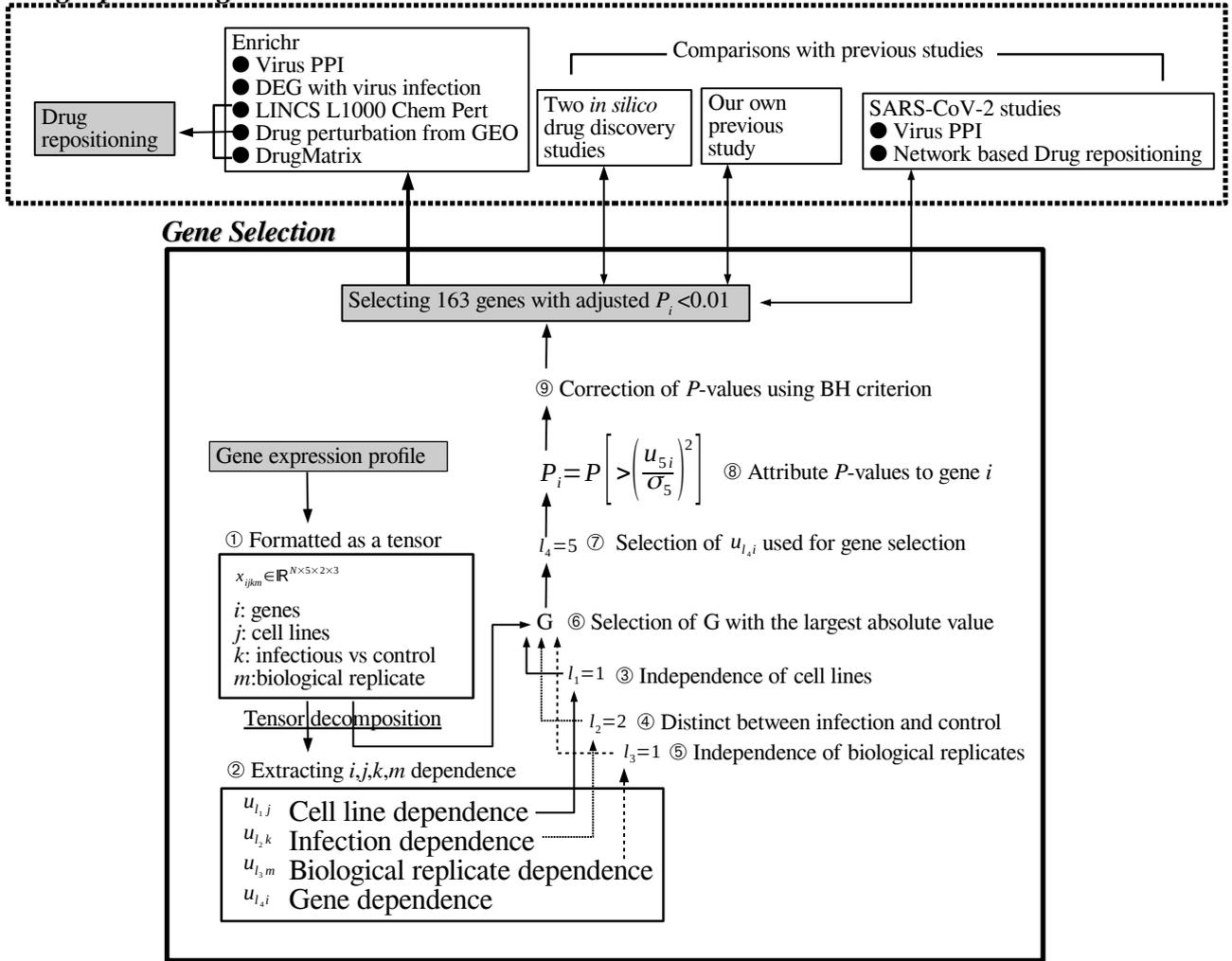


図 1 この解析の全体の流れ。

Fig. 1 Overall design of this study.

染の2通りの処理を3回ずつ行った (Biological replicate) 計30サンプルからなっている遺伝子発現プロファイルである。

2.2 テンソル分解を用いた教師なし学習による変数選択法
遺伝子発現プロファイルは、

$$x_{ijkm} \in \mathbb{R}^{N \times 5 \times 2 \times 3} \quad (1)$$

のテンソルの形式に変換される。 $N = 21,797$ は総遺伝子数である。 x_{ijkm} は i 番目の遺伝子の、 j 番目の培養細胞の ($j = 1$: Calu3, $j = 2$: NHBE, $j = 3$: A549 MOI 0.2, $j = 4$: A549 MOI 2.0, $j = 5$: A549 ACE2 expressed) k 番目の処理 ($k = 1$: モック、及び $k = 2$: SARS-CoV-2 の感染) の m 個目の Biological Replicate における発現量を示す。 x_{ijkm} は

$$\sum_i x_{ijkm} = 0 \quad (2)$$

$$\sum_i x_{ijkm}^2 = N \quad (3)$$

となるように規格化されている。

x_{ijkm} に HOSVD (Higher Order Singular Value Decomposition) [3] を適用し、以下を得る。

$$x_{ijkm} = \sum_{l_1=1}^5 \sum_{l_2=1}^2 \sum_{l_3=1}^3 \sum_{l_4=1}^N G(l_1 l_2 l_3 l_4) \times u_{l_1 j} u_{l_2 k} u_{l_3 m} u_{l_4 i} \quad (4)$$

ここで $u_{l_1 j} \in \mathbb{R}^{5 \times 5}$, $u_{l_2 k} \in \mathbb{R}^{2 \times 2}$, $u_{l_3 m} \in \mathbb{R}^{3 \times 3}$, $u_{l_4 i} \in \mathbb{R}^{N \times N}$ は特異値行列であり直交行列である。 $G(l_1 l_2 l_3 l_4) \in \mathbb{R}^{5 \times 2 \times 3 \times N}$ はコアテンソルで、 l_1, l_2, l_3, l_4 の組み合わせの寄与の大きさを表す。

以下を満たす l_1, l_2, l_3 を探す

- $u_{l_1 j}$ に j 依存性がない (培養細胞の種類に依らない発現量)。
- $u_{l_2 1}$ と $u_{l_2 2}$ が逆符号 (感染の有無で差がある)。

- $u_{\ell_3 m}$ に m 依存性がない (Biological Replicate に依らない発現量)。

この様な l_1, l_2, l_3 が求まったら、 l_1, l_2, l_3 を固定した上で、絶対値が最大の $G(l_1 l_2 l_3 l_4)$ を与える l_4 を見つける。

l_4 が見つかったら、 $u_{\ell_4 i}$ がガウス分布であることを仮定し (帰無仮説)、

$$P_i = P_{\chi^2} \left[> \left(\frac{u_{\ell_4 i}}{\sigma_{\ell_4}} \right)^2 \right] \quad (5)$$

で i 番目の遺伝子に P 値を付与する。ここで $P_{\chi^2}[> x]$ は引数が x 以上である場合の累積 χ^2 乗分布、 σ_{ℓ_4} は $u_{\ell_4 i}$ の標準偏差である。

P_i は BH 基準 [3] で多重比較補正され、補正 P 値が 0.01 以下の遺伝子 i を条件を満たすものとして選択する。

3. 結果

3.1 遺伝子選択

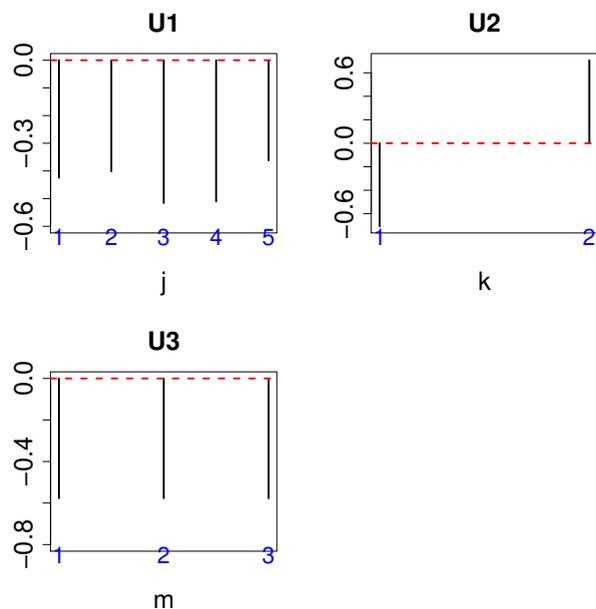


図 2 HOSVD で求めた特異値ベクトル。U1: u_{1j} , U2: u_{2k} , U3: u_{1m} , j, k, m の意味は方法を見よ。

Fig. 2 Singular value vectors obtained by the HOSVD algorithm. U1: u_{1j} , U2: u_{2k} , U3: u_{1m} , See Materials and methods for the definitions of j, k , and m .

図 2 に u_{1j}, u_{2k}, u_{1m} を示す。この図から $l_1 = l_3 = 1, l_2 = 2$ が要請を満たすと解る。そこで $G(1 2 1 l_4)$ の絶対値が最大となる l_4 を探す。表 1 にあるようにこのような l_4 は $l_4 = 5$ であることが解る。そこで (5) 式に $l_4 = 5$ を代入して P_i を遺伝子 i に付与した。補正 P 値が 0.01 以下の遺伝子を選んだ結果 163 個の遺伝子が選択された (表 2)。

3.2 エンリッチメント解析

前小節で選択した 163 遺伝子を Enrichr [4] にアップ

表 1 HOSVD で計算された $G(1 2 1 l_4)$

Table 1 $G(1 2 1 l_4)$ s computed by the HOSVD algorithm

| l_4 | $G(1 2 1 l_4)$ | l_4 | $G(1, 2, 1, l_4)$ |
|-------|----------------|-------|-------------------|
| 1 | -21.409671 | 6 | -12.388615 |
| 2 | 5.183297 | 7 | 8.437642 |
| 3 | -21.426437 | 8 | 13.322888 |
| 4 | 10.030564 | 9 | -1.850982 |
| 5 | 62.518121 | 10 | 9.211437 |

表 2 テンソル分解を用いた教師なし学習による変数選択法で選ばれた 163 遺伝子

Table 2 One hundred and sixty-three genes selected by TD-based unsupervised FE

ABCC3 ACE2 ACTB ACTG1 ACTN4 AHNAK AKAP12
AKR1B1 AKR1B10 AKR1C2 ALDH1A1 ALDH3A1 ALDOA
AMIGO2 ANTXR1 ANXA2 ASNS ASPH ATF4 ATP1B1 C3
CALM2 CALR CD24 CFL1 CPLX2 CRIM1 CTGF CXCL5
CYP24A1 DCBLD2 DDIT4 DHCR24 EEF1A1 EEF2 EIF1
EIF4B EIF5A ENO1 ERBB2 EREG FADS2 FASN FDCSP
FDPS FLNB FTH1 FTL G6PD GAPDH GAS5 GPX2 GSTP1
H1FO HMGA1 HNRNPA2B1 HSP90AA1 HSP90AB1 HSPA8
ICAM1 IER3 IFIT2 IGFBP3 IGFBP4 ITGA2 ITGA3 ITGAV
ITGB1 JUN KRT18 KRT19 KRT23 KRT5 KRT6A KRT7
KRT8 KRT81 LAMB3 LAMC2 LCN2 LDHA LIF LOXL2
MIEN1 MTHFD2 MYL6 NAMPT NAP1L1 NEAT1 NFK-
BIA NPM1 NQO1 OAS2 P4HB PABPC1 PFN1 PGK1 PKM
PLAU PLOD2 PMEPA1 PPIA PPP1R15A PSAT1 PSMD3
PTMA RAI14 RNF213 RPL10 RPL12 RPL23 RPL26 RPL28
RPL3 RPL37 RPL4 RPL5 RPL7 RPL7A RPL9 RPS19 RPS20
RPS24 RPS27 RPS27A RPS3A RPS4X RPS6 S100A2 S100A6
SAT1 SCD SERPINA3 SERPINE1 SLC38A2 SLC7A11 SLC7A5
SPP1 SPTBN1 SQSTM1 STARD3 STAT1 STC2 TGFBI TGM2
TIPARP TMSB4X TNFAIP2 TOP2A TPI1 TPM1 TPT1
TRAM1 TUBA1B TUBB TUBB4B TXNIP TXNRD1 UBC
VEGFA VIM YBX1 YWHAZ

ロードしてエンリッチメント解析をすることで生物学的な妥当性を検証した。論文を書いた当時は SARS-CoV-2 の情報は殆ど無かったため、近縁の SARS の情報で検証した。表 3 は SARS のウイルスタンパクの内、選択された 163 個のヒト遺伝子と相互作用するタンパクのリストである。非常に多数の SARS ウイルスタンパクが、選択された 163 遺伝子と相互作用しているため、このリストは非常に有望であると思われる。

Enrichr には SARS-CoV-2 の情報は当時は無かったのでこれ以上検証できなかったが、SARS-CoV-2 のデータを別途取ってきてやってみたのが表 4 である。明らかに SARS の時より有意性が劇的に向上している。この様に選択された 163 遺伝子は生物学的に妥当な様に思われた。他にも評価は行ったので原著論文 [1] を見ていただきたい。

表 3 ウィルスタンパクの内、テンソル分解を用いた教師なし学習により選択された 163 遺伝子と有意に相互作用しているタンパク。Enrichr の “Virus-Host PPI P-HIPSTER 2020” カテゴリーに基づく

Table 3 Virus proteins that significantly interact with the 163 genes selected by TD based unsupervised FE and enriched by “Virus-Host PPI P-HIPSTER 2020” in Enrichr

| Term | Overlap | P-value | Adjusted P-value |
|--|---------|------------------------|------------------------|
| SARS coronavirus excised_polyprotein 1..4369 (gene: orf1ab) | 12/194 | 6.67×10^{-8} | 2.38×10^{-6} |
| SARS coronavirus P2_full_polyprotein 1..4382 | 12/198 | 8.35×10^{-8} | 2.76×10^{-6} |
| SARS coronavirus hypothetical protein sars9b | 4/17 | 9.31×10^{-6} | 7.57×10^{-5} |
| SARS coronavirus P2 hypothetical protein sars9b | 4/17 | 9.31×10^{-6} | 7.562×10^{-5} |
| SARS coronavirus Tor2 Orf13 | 4/17 | 9.31×10^{-6} | 7.55×10^{-5} |
| SARS coronavirus nsp7-pp1a/pp1ab (gene: orf1ab) | 5/36 | 1.038×10^{-5} | 8.18×10^{-5} |
| SARS coronavirus 3C-like proteinase (gene: orf1ab) | 4/19 | 1.49×10^{-5} | 1.10×10^{-4} |
| SARS coronavirus nucleocapsid protein (gene: N) | 4/29 | 8.61×10^{-5} | 4.23×10^{-4} |
| SARS coronavirus P2 nucleocapsid protein | 4/29 | 8.61×10^{-5} | 4.23×10^{-4} |
| SARS coronavirus Tor2 nucleocapsid protein | 4/29 | 8.61×10^{-5} | 4.22×10^{-4} |
| SARS coronavirus nsp4-pp1a/pp1ab (gene: orf1ab) | 3/16 | 2.75×10^{-4} | 9.89×10^{-4} |
| SARS coronavirus formerly known as growth-factor-like protein (gene: orf1ab) | 3/17 | 3.32×10^{-4} | 1.14×10^{-3} |
| SARS coronavirus nsp8-pp1a/pp1ab (gene: orf1ab) | 4/45 | 4.88×10^{-4} | 1.50×10^{-3} |
| SARS coronavirus leader protein (gene: orf1ab) | 3/20 | 5.47×10^{-4} | 1.61×10^{-3} |
| SARS coronavirus RNA-dependent RNA polymerase (gene: orf1ab) | 2/9 | 2.28×10^{-3} | 5.26×10^{-3} |
| SARS coronavirus P2 spike glycoprotein precursor | 4/71 | 2.70×10^{-3} | 6.08×10^{-3} |
| SARS coronavirus nsp3-pp1a/pp1ab (gene: orf1ab) | 5/118 | 2.82×10^{-3} | 6.34×10^{-3} |
| SARS coronavirus E2 glycoprotein precursor (gene: S) | 4/72 | 2.84×10^{-3} | 6.38×10^{-3} |
| SARS coronavirus Tor2 spike glycoprotein | 4/72 | 2.84×10^{-3} | 6.38×10^{-3} |
| SARS coronavirus 2-O-ribose methyltransferase (2-o-MT) (gene: orf1ab) | 2/11 | 3.45×10^{-3} | 7.26×10^{-3} |
| SARS coronavirus hypothetical protein sars7a | 3/38 | 3.63×10^{-3} | 7.59×10^{-3} |
| SARS coronavirus P2 hypothetical protein sars7a | 3/38 | 3.63×10^{-3} | 7.58×10^{-3} |
| SARS coronavirus Tor2 Orf8 | 3/38 | 3.63×10^{-3} | 7.58×10^{-3} |
| SARS coronavirus nsp9-pp1a/pp1ab (gene: orf1ab) | 2/13 | 4.85×10^{-3} | 9.45×10^{-3} |
| SARS coronavirus nsp13-pp1ab (ZD, NTPase/HEL; RNA) (gene: orf1ab) | 2/14 | 5.63×10^{-3} | 1.06×10^{-2} |
| SARS coronavirus Tor2 replicase 1AB | 4/108 | 1.18×10^{-2} | 1.94×10^{-2} |
| SARS coronavirus P2_full_polyprotein 1..7073 | 4/109 | 1.22×10^{-2} | 2.00×10^{-2} |

3.3 ドラッグディスカバリー

次にこの 163 遺伝子の発現量に、投与が有意に影響を与えることが解っている既存薬をスクリーニングすることでドラッグリポジショニングを行った (表 5、Enrichr による)。複数のデータセットでヒットがあっただけではなく、既報 [5] の候補薬剤リストともよく一致していた。このことから、提案手法は有効な手法であると思われる (ドラッグリポジショニングの詳細は原著論文を参照 [1] のこと)。

4. 議論

他の遺伝子選択手法との比較を行った (表 6)。どの方法も、培養細胞種に依らない一定の DEG の選択に失敗している。特に DESeq2 は業界の SOTA とされているにもかかわらず、細胞種ごとに極端に数の大小がある DEG 選択となってしまうことが解る。そもそも、全体で 30 サンプルしかないから、5 種類の培養細胞ごとに

別々に DEG を選択したら 6 サンプルしか無いわけであり、安定的な選択など望むべくもない。

提案手法が DEG の選択に於いて圧倒的に優れていることは明らかである。

5. 結論

本研究報告では、テンソル分解を用いた教師なし学習による変数選択法を SARS-CoV-2 に感染したヒト肺培養細胞の遺伝子発現プロファイルに適用することで、163 個のヒト遺伝子を選ぶことに成功した。これらの遺伝子をエンリッチメント解析で評価した所、生物学的妥当な結果だと思われたため、その遺伝子の発現量に、投与が有意に影響を与えることが解っている既存薬をドラッグリポジショニングした所、既報とよく合っていたので有望だと思われた。最後に遺伝子の選択自体を他の既存手法と比較したが妥当な数の DEG を選択するのはテンソル分解を用いた教

表 4 テンソル分解を用いた教師なし学習により選択された163遺伝子と有意に相互作用している SARS-CoV-2 タンパク

Table 4 Coincidence between 163 genes and human proteins known to interact with SARS-CoV-2 proteins.

| SARS-CoV-2 proteins | P values | Odds Ratio |
|----------------------|------------------------|------------|
| SARS-CoV2 E | 6.55×10^{-27} | 10.16 |
| SARS-CoV2 M | 1.38×10^{-26} | 8.42 |
| SARS-CoV2 N | 4.61×10^{-24} | 11.43 |
| SARS-CoV2 nsp1 | 1.06×10^{-20} | 10.00 |
| SARS-CoV2 nsp10 | 3.40×10^{-20} | 11.52 |
| SARS-CoV2 nsp11 | 1.13×10^{-29} | 10.66 |
| SARS-CoV2 nsp12 | 4.87×10^{-20} | 9.48 |
| SARS-CoV2 nsp13 | 6.04×10^{-33} | 11.17 |
| SARS-CoV2 nsp14 | 1.75×10^{-22} | 12.05 |
| SARS-CoV2 nsp15 | 1.85×10^{-20} | 10.23 |
| SARS-CoV2 nsp2 | 4.81×10^{-33} | 11.79 |
| SARS-CoV2 nsp4 | 5.79×10^{-29} | 10.26 |
| SARS-CoV2 nsp5 | 3.78×10^{-25} | 12.36 |
| SARS-CoV2 nsp5_C145A | 3.75×10^{-17} | 11.39 |
| SARS-CoV2 nsp6 | 9.47×10^{-26} | 9.00 |
| SARS-CoV2 nsp7 | 1.93×10^{-29} | 10.81 |
| SARS-CoV2 nsp8 | 1.11×10^{-29} | 10.14 |
| SARS-CoV2 nsp9 | 5.54×10^{-29} | 12.24 |
| SARS-CoV2 orf10 | 5.29×10^{-34} | 12.37 |
| SARS-CoV2 orf3a | 2.06×10^{-28} | 9.95 |
| SARS-CoV2 orf3b | 1.89×10^{-29} | 11.80 |
| SARS-CoV2 orf6 | 8.81×10^{-26} | 10.37 |
| SARS-CoV2 orf7a | 1.69×10^{-28} | 10.00 |
| SARS-CoV2 orf8 | 5.94×10^{-28} | 9.25 |
| SARS-CoV2 orf9b | 6.54×10^{-30} | 12.12 |
| SARS-CoV2 orf9c | 1.11×10^{-28} | 8.35 |
| SARS-CoV2 Spike | 8.22×10^{-26} | 10.08 |

教師なし学習による変数選択法だけであり、方法論的な優位性が確保できた。

参考文献

- [1] Taguchi, Y.-h. and Turki, T.: A new advanced in silico drug discovery method for novel coronavirus (SARS-CoV-2) with tensor decomposition-based unsupervised feature extraction, *PLOS ONE*, Vol. 15, No. 9, pp. 1–16 (online), DOI: 10.1371/journal.pone.0238907 (2020).
- [2] Tsuji, M.: Potential anti-SARS-CoV-2 drug candidates identified through virtual screening of the ChEMBL database for compounds that target the main coronavirus protease, *FEBS Open Bio*, Vol. 10, No. 6, pp. 995–1004 (online), DOI: <https://doi.org/10.1002/2211-5463.12875> (2020).
- [3] Taguchi, Y.-H.: *Unsupervised Feature Extraction Applied to Bioinformatics*, Springer International Publishing (2020).
- [4] Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W. and Ma'ayan, A.: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, *Nucleic Acids Re-*

表 5 選択された既存薬と、既報 [5] で報告された化合物のリストとの一致

Table 5 Number of experiments associated with adjusted *P*-values in various Enrichr categories for the drugs identified in another study [5]

| | Drug Matrix | GEO | | LINCS | |
|--------------------|-------------|------|----|-------|----|
| | | down | up | down | up |
| Mestranol | 6 | | | | |
| Methotrexate | 40 | 2 | | | |
| Fluorouracil | 4 | 4 | 6 | 1 | |
| Hexestrol | 4 | | | | |
| Mercaptopurine | 17 | | | | |
| Paroxetine | 10 | | | 1 | |
| Vinblastine | 17 | | | 1 | 3 |
| Phenylbutazone | 6 | | | | |
| Naloxone | 5 | | | | |
| Hydralazine | 14 | | | | |
| Vinorelbine | 13 | | | | 1 |
| Carvedilol | 20 | | | | 1 |
| Colchicine | 13 | | 1 | | |
| Amitriptyline | 13 | | | 1 | |
| Epinephrine | 17 | | | | |
| Dactinomycin | 6 | 2 | 1 | 5 | 1 |
| Hydrocortisone | 21 | | | | |
| Melatonin | 7 | | | | |
| Methyltestosterone | 6 | | | | |
| Omeprazole | 20 | | | | |
| Testosterone | 12 | | 1 | | |
| Oxymetholone | 5 | | | | |
| Progesterone | 20 | 2 | 2 | | |
| Permethrin | | 1 | | | |
| Mesalazine | | 1 | | | |
| Menadione | | 1 | 1 | | 3 |
| Stanolone | | 1 | 2 | | |
| Methotrexate | | | 1 | | |
| Sirolimus | | | | 47 | 50 |
| Cladribine | | | | 2 | 1 |
| Regorafenib | | | | 9 | 14 |
| Temsirolimus | | | | 1 | |
| Quinacrine | | | | 1 | 3 |

search, Vol. 44, No. W1, pp. W90–W97 (online), DOI: 10.1093/nar/gkw377 (2016).

- [5] Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W. and Cheng, F.: Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2, *Cell Discovery*, Vol. 6, No. 1 (online), DOI: 10.1038/s41421-020-0153-3 (2020).
- [6] Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W.-C., Uhl, S., Hoagland, D., Mller, R., Jordan, T. X., Oishi, K., Panis, M., Sachs, D., Wang, T. T., Schwartz, R. E., Lim, J. K., Albrecht, R. A. and tenOever, B. R.: Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19, *Cell*, Vol. 181, No. 5, pp. 1036–1045.e9 (online), DOI: <https://doi.org/10.1016/j.cell.2020.04.026> (2020).
- [7] Tusher, V. G., Tibshirani, R. and Chu, G.: Significance

表 6 DEG の比較。 t 検定、SAM、limma、DESeq2。 DESeq2 の結果はだけは原論文 [6] より

Table 6 DEG identifications between control and infectious cell lines using t test, SAM, limma and DESeq2. Genes associated with adjusted P -values less than 0.01 are selected as DEG.

| | t test | | SAM [7] | | limma [8] | | DESeq2 [9] | |
|----------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|
| | $P \geq 0.01$ | $P < 0.01$ |
| Calu3 | 21754 | 43 | 21797 | 0 | 42 | 13380 | 7278 | 16432 |
| NHBE | 21797 | 0 | 21797 | 0 | 41 | 13328 | 23383 | 327 |
| A549 | | | | | | | | |
| MOI 0.2 | 21797 | 0 | 21797 | 0 | 50 | 13867 | 7858 | 15852 |
| MOI 2.0 | 21472 | 325 | 21797 | 0 | 15 | 13823 | 16279 | 7431 |
| ACE2 expressed | 21796 | 1 | 21797 | 0 | 111 | 11403 | 16201 | 7509 |

analysis of microarrays applied to the ionizing radiation response, *Proceedings of the National Academy of Sciences*, Vol. 98, No. 9, pp. 5116–5121 (online), DOI: 10.1073/pnas.091062498 (2001).

- [8] Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K.: limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Research*, Vol. 43, No. 7, pp. e47–e47 (online), DOI: 10.1093/nar/gkv007 (2015).
- [9] Love, M. I., Huber, W. and Anders, S.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biology*, Vol. 15, No. 12 (online), DOI: 10.1186/s13059-014-0550-8 (2014).