 doi:10.20944/preprints202004.0524.v1

Highlights

A New Advanced *In silico* Drug Discovery Method for Novel Coronavirus (SARS-CoV-2) with Tensor Decomposition-based Unsupervised Feature Extraction

Y-h. Taguchi,Turki Turki

- We performed *in-silico drug* discovery for SARS-CoV-2.
- We propose unsupervised learning approach with a new tensor decomposition formalism.
- We identified drug candidate compounds and 163 important genes in SARS-CoV-2.
- This study contributes to advancing drug screening in COVID-19 infectious diseases.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 April 2020 doi:10.20944/preprints202004.0524.v1

A New Advanced *In silico* Drug Discovery Method for Novel Coronavirus (SARS-CoV-2) with Tensor Decomposition-based Unsupervised Feature Extraction

Y-h. Taguchi[a,*], Turki Turki[b]

[a]Department of Physics, Chuo University, 112-8551 Tokyo, Japan
[b]King Abdulaziz University, Department of Computer Science, Jeddah, 21589, Saudi Arabia

ARTICLE INFO

Keywords:
unsupervised learning
tensor decomposition
feature selection
COVID-19
drug discovery
gene expression

ABSTRACT

Background: COVID-19 is a critical pandemic that has affected human communities worldwide. Although it is urgent to rapidly develop effective drugs, large number of candidate drug compounds may be useful for treating COVID-19, and evaluation of these drugs is time-consuming and costly. Thus, screening to identify potentially effective drugs prior to experimental validation is necessary. Method: In this study, we applied the recently proposed method tensor decomposition (TD)-based unsupervised feature extraction (FE) to gene expression profiles of multiple lung cancer cell lines infected with severe acute respiratory syndrome coronavirus 2. We identified drug candidate compounds that significantly altered the expression of the 163 genes selected by TD-based unsupervised FE. Results: Numerous drugs were successfully screened, including many known antiviral drug compounds. Conclusions: The drugs screened using our strategy may be effective candidates for treating patients with COVID-19.

1. Introduction

Coronavirus 2019 (COVID-19) is an infectious disease that has created a pandemic worldwide [18]. Thus, it is urgent to identify effective drugs to combat this disease. Numerous studies related to identifying effective therapeutics have been reported; *in slico* drug discovery is a useful approach because very large numbers (up to millions) of drug candidate compounds can be screened, which is not possible using experimental approaches. There are two main methods used for *in slico* drug discovery, ligand-based drug discovery (LBDD) and structure-based drug discovery (SBDD), which have various advantages and disadvantages. LBDD can effectively predict "hit" compounds but cannot find new drug candidate compounds lacking similarity to known drug compounds. In contrast, although SBDD can find drug candidate compounds without similarity to known drugs, it requires massive computational resources for docking simulation between compounds and proteins. When no experimentally confirmed protein tertiary structures are available, these structures must also be predicted, potentially decreasing the accuracy of the predicted affinity of compounds with proteins. If gene expression profiles altered by new drug candidate compounds are coincident with those of known drug compounds, these new drug candidate compounds are regarded as promising. Although this approach can identify promising drug candidate compounds even when they lack similarity with known drugs, as required by LBDD, and massive computational resources are not needed, as required by SBDD, it remains difficult to identify drug candidate compounds for proteins and diseases when no effective drug compounds are known.

To overcome these limitations, we propose an unsupervised method that can predict drug candidate compounds without knowledge of known compounds using a different formulation of the recently proposed tensor decomposition (TD)-based unsupervised feature extraction (FE) [24, 22, 26, 23]. TD-based unsupervised FE was applied to the gene expression profiles of multiple lung cancer cell lines infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [2]. The 163 genes identified as differentially expressed genes (DEGs) in SARS-CoV-2 infection were enriched in various SARS coronavirus-related terms. Drugs screened based on the coincidence of DEGs between drug treatments and SARS-CoV-2 infection were largely enriched with known antivirus drugs. This suggests that our strategy is effective and that the drugs screened in this study are promising candidates as antiviral drug for SARS-CoV-2.

2. Materials and Methods

2.1. Gene expression profiles

Gene expression profiles used in this study were downloaded from the Gene Expression Omnibus (GEO) with GEO ID GSE147507. It is composed of five cell lines (Calu3, NHBE, A549 Multiplicity of infection (MOI) 0.2, A549 MOI 2,0, and A549 ACE2 expressed), two treatments (Mock and SARS-CoV-2 infected), and three biological replicates for individual pairs of cell lines and treatments. Thus, in total, $5 \times 2 \times 3 = 30$ samples were available.

*This work was supported by KAKENHI [grant numbers 19H05270, 20H04848, and 20K12067] and Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah [grant number KEP-8-611-38]. The authors, therefore, acknowledge DSR with thanks for providing technical and financial support.

*Corresponding author
tag@granular.com (Y-h. Taguchi); tturki@kau.edu.sa (T. Turki)
ORCID(s): 0000-0003-0867-8986 (Y-h. Taguchi); 0000-0002-9491-2435 (T. Turki)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 April 2020 doi:10.20944/preprints202004.0524.v1

2.2. TD-based unsupervised FE

Gene expression profiles are formatted as tensor, $x_{ijkm} \in \mathbb{R}^{N\times5\times2\times3}$, which represents the ith gene expression of jth cell lines ($j = 1$: Calu3, $j = 2$: NHBE, $j = 3$: A549 MOI 0.2, $j = 4$: A549 MOI 2,0, $j = 5$: A549 ACE2 expressed) with kth treatment ($k = 1$: Mock and $k = 2$: SARS-CoV-2 infected) of the mth biological replicates.

x_{ijkm} was decomposed into TD

$$x_{ijkm} = \sum_{\ell_1=1}^{5}\sum_{\ell_2=1}^{2}\sum_{\ell_3=1}^{3}\sum_{\ell_4=1}^{N} G(\ell_1,\ell_2,\ell_3,\ell_4,\ell_5)u_{\ell_1 j}u_{\ell_2 k}u_{\ell_3 m}u_{\ell_4 i} \tag{1}$$

with a higher-order singular value decomposition (HOSVD) [23]. $u_{\ell_1 j} \in \mathbb{R}^{5\times5}, u_{\ell_2 k} \in \mathbb{R}^{2\times2}, u_{\ell_3 m} \in \mathbb{R}^{3\times3}, u_{\ell_4 i} \in \mathbb{R}^{N\times N}$ are singular value matrices which are orthogonal matrices. The tensor was normalized as $\sum_i x_{ijkm} = 0$ and $\sum_i x^2_{ijkm} = N$. $G(\ell_1,\ell_2,\ell_3,\ell_4) \in \mathbb{R}^{N\times5\times2\times3}$ is a core tensor that represents a weight of the combination of $\ell_1,\ell_2,\ell_3,\ell_4$

To identify $u_{\ell_4 i}$ which is used for gene selection, we need to identify $u_{\ell_1 j}$ whose values are independent of j, i.e. cell line-independent, $u_{\ell_2 m}$ whose values are independent of m, i.e., biological replicate-independent while $u_{\ell_2 k}$ whose values are distinct between $k = 1$ and $k = 2$, i.e., distinct between Mock infection and SARS-CoV-2.

The next step was to identify $G(\ell_1,\ell_2,\ell_3,\ell_4)$ with the largest absolute values with fixed ℓ_1,ℓ_2,ℓ_3. This enabled selection of $u_{\ell_4 i}$ used for gene selection. P-values, P_is, are attributed to ith gene using the following formula:

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

where $P_{\chi^2}[> x]$ is cumulative distribution of the χ^2 distribution where the argument is larger than x. Next, P_is was adjusted by Benjamini and Hochberg criterion [23] and genes associated with adjusted P-values less than 0.01 were selected.

2.3. Enrichment analysis

Gene symbols of genes selected by TD-based unsupervised FE with significantly altered expression due to SARS-CoV-2 infection were uploaded to Enricher [14], which is a popular enrichment analysis server that evaluates the biological properties of genes based on enrichment analysis.

3. Results

3.1. Gene selection

After identifying $\ell_1 = 1, \ell_2 = 2$, and $\ell_3 = 1$ based upon the criterion denoted in the Materials and Methods (Fig. 1), we attempted to list $G(1,2,1,\ell_4)$s to select ℓ_4 used for gene selection. We found that $G(1,2,1,5)$ had the largest absolute value (Table 1). As a result, u_{5i} was employed to attribute P-values to gene i as shown in Eq. (2). Finally, we selected 163 genes showing adjusted P-values less than 0.01 (Table 2).

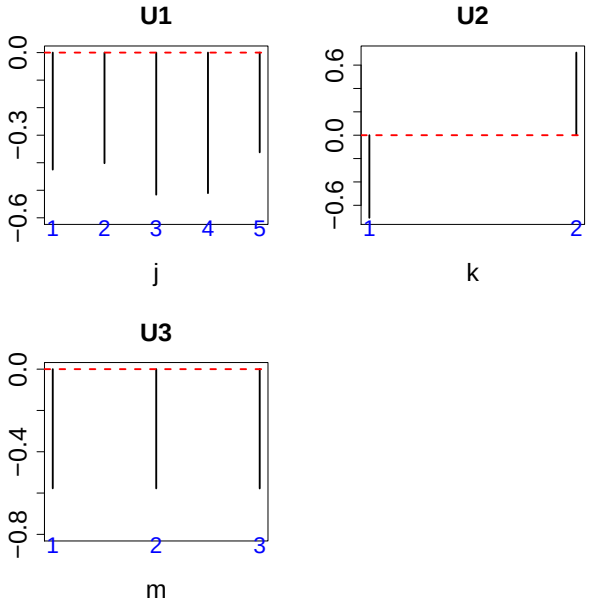

Figure 1: Singular value vectors obtained by the HOSVD algorithm. U1:U_{1j}, U2:U_{2k}, U3:U_{1m}, See Materials and Methods for the definitions of j, k, and m.

Table 1
$G(1,2,1,\ell_4)$s computed by the HOSVD algorithm

ℓ_4	$G(1,2,1,\ell_4)$	ℓ_4	$G(1,2,1,\ell_4)$
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

3.2. Enrichment analysis

The selected 163 genes were uploaded to Enrichr (full list is available in the supplementary materials) and we identified numerous enriched categories useful for follow-up analyses of the selected 163 genes and in *in silico* drug discovery as described below.

3.2.1. Protein-protein interactions

The 163 selected proteins significantly interacted with numerous SARS-CoV virus proteins with critical roles in virus infection. Thus, our strategy can successfully identify critical human genes during coronavirus infection (Table 3, full list is available in the supplementary materials).

3.2.2. Virus perturbations

Next, we examined whether the selected 163 genes significantly overlapped with genes whose expression was altered by infection with viruses other than SARS-CoV-2. We investigated "Virus Perturbations from GEO up" (Table 4, full list is available in the supplementary materials) and "Virus Perturbations from GEO down" (Table 5, full list is available in the supplementary materials). We found that SARS-CoV and SARS-BAtSRBD, which are coronaviruses mostly re-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 April 2020 doi:10.20944/preprints202004.0524.v1

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 H1F0 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 NFKBIA 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

lated to SARS-CoV-2, were highly enriched. This also suggests that our strategy is effective for identifying genes important in SARS-CoV-2 infection.

3.3. Drug discovery

Based upon the observations described above, we regarded the selected 163 proteins as representative of the SARS-CoV-2 infection process. Next, we evaluated drug candidate compounds by identifying those that significantly affected the expression of the selected 163 genes. For this, we investigated individual drug treatment-related categories in Enrichr.

3.3.1. LINCS L1000 Chem Pert up/down

The first category investigated in Enrichr was "LINCS L1000 chem pert". LINCS collected numerous cell lines treated with various drug compounds. Their altered expression profiles have been measured and stored in a public domain database. We found many drug compounds whose treatments significantly altered the expression of the selected 163 genes. Because the number of "hits" is too large to show here, tables are provided as supplementary information. Selected drugs in this category are shown below. We identified many candidate drug compounds, indicating that our strategy is effective.

C646 C646 showed the second smallest (significant) *P*-value in "LINCS L1000 Chem Pert up" and had multiple hits (Table S1). This agent was also reported to be a novel p300/CREB-binding protein-specific inhibitor of histone acetyltransferase which attenuates influenza A virus infection [33].

Chelerythrine chloride Chelerythrine chloride exhibited the third and fifth smallest (significant) *P*-value in "LINCS L1000 Chem Pert up" and had multiple hits (Table S2). It is known to exhibit pharmacological inhibition of protein kinase C reduces West Nile virus replication (See Fig,1 [3]).

Canertinib Canertinib exhibited the sixth smallest (significant) *P*-value in "LINCS L1000 Chem Pert up" and had multiple hits (Tables S3 and S4). It shows antiviral chemotherapy effects and controls poxvirus infections by inhibiting cellular signal transduction [31].

BX-795 BX-795 has the 11th smallest (significant) *P*-value in "LINCS L1000 Chem Pert up" and had multiple hits (Table S5). BX-795 inhibits HSV-1 and HSV-2 replication by blocking the JNK/p38 pathways without interfering with PDK1 activity in host cells [20]. Su et al [20] also suggested SARS-CoV as a target of BX-795.

Sorafenib Sorafenib showed the 12th smallest (significant) *P*-value in "LINCS L1000 Chem Pert up" and had multiple hits (Table S6). Sorafenib impedes Rift Valley fever virus egress by inhibiting valosin-containing protein function in the cellular secretory pathway [4].

QL-X-138 QL-X-138 displayed the smallest (significant) *P*-value in "LINCS L1000 Chem Pert down" and had multiple hits (Tables S7 and S8). QL-XII-138 inhibits Dengue virus (see Figure 3 [28]).

Radicicol Radicicol showed the second smallest (significant) *P*-value in "LINCS L1000 Chem Pert down" and had multiple hits (Tables S9 and S10). Antiviral activity and RNA polymerase of radicicol is degradation following Hsp90 inhibition in a range of negative-strand viruses [6]. Radicicol also preferentially reduces HCV release, although radicicol does not affect its infectivity [13]. Because other Hsp90 inhibitors are effective against coronavirus [15], radicidol is also thought to be effective for treating SARS-CoV-2.

A-443654 A-443654 shoewd the fourth smallest (significant) *P*-value in "LINCS L1000 Chem Pert down" and had multiple hits (Tables S11 and S12). Jeong and Ahn found that viral replication of HBV in infected or transfected hepatoma cells was markedly inhibited by treatment with A-443654 [12], a specific inhibitor of Akt. As the SARS-CoV membrane protein also induces apoptosis by modulating the Akt survival pathway [5], A-443654 may be an effective drug for treating COVID-19. The "PI3K-Akt signaling pathway" was the fourth most significant pathway (adjusted $P = 3.97\times 10^{-7}$, overlap is 17/354) in the "KEGG 2019 Human" category of Enrichr (full list is available in the supplementary materials) to which the 163 selected genes were uploaded.

CGP-60474 CGP-60474 had the fifth smallest (significant) *P*-value in "LINCS L1000 Chem Pert down" and multiple hits (Tables S13 and S14). CGP-60474 is also a repurposed drug which was used to treat lung injury in COVID-19 in an independent *in silico* study [11].

Table 3
Virus proteins that significantly interact with the 163 genes selected by TD based unsupervised FE due to "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}

Table 4
Genes whose expression is altered by SARS-CoV-2 related viruses that significantly interact with the 163 genes selected by TD based unsupervised FE due to "Virus Perturbations from GEO up" in Enrichr

Term	Overlap	P-value	Adjusted P-value
SARS-BatSRBD 48Hour GSE47960	11/300	3.66×10^{-5}	1.48×10^{-3}
SARS-CoV 12Hour GSE17400	11/300	3.66×10^{-5}	1.31×10^{-3}
SARS-CoV 48Hour GSE47961	11/300	3.66×10^{-5}	1.18×10^{-3}
icSARS CoV 54Hour GSE37827	11/300	3.66×10^{-5}	1.08×10^{-3}
SARS-CoV MA15 Day2 GSE49263	10/300	1.82×10^{-4}	3.45×10^{-3}
SARS-CoV 60Hour GSE47960	10/300	1.82×10^{-4}	3.26×10^{-3}
SARS-CoV 96Hour GSE47961	10/300	1.82×10^{-4}	3.09×10^{-3}
SARS-ddORF6 24Hour GSE47961	10/300	1.82×10^{-4}	2.93×10^{-3}
SARS-BatSRBD 36Hour GSE47960	9/300	8.14×10^{-4}	1.05×10^{-2}
SARS-CoV MA15 Day4-C57BL-6 GSE40824	9/300	8.14×10^{-4}	1.01×10^{-2}
SARS-dORF6 72Hour GSE47960	9/300	8.14×10^{-4}	9.73×10^{-3}
SARS-ddORF6 72Hour GSE47961	9/300	8.14×10^{-4}	9.39×10^{-3}
SARS-BatSRBD 84Hour GSE47961	8/300	3.27×10^{-3}	2.46×10^{-2}
SARS-CoV MA15 Day2 GSE49262	8/300	3.27×10^{-3}	2.40×10^{-2}
SARS-CoV MA15 Day4-PFU-10^5 GSE33266	8/300	3.27×10^{-3}	2.35×10^{-2}
SARS-dORF6 84Hour GSE47962	8/300	3.27×10^{-3}	2.30×10^{-2}
cSARS Bat SRBD 24Hour GSE37827	8/300	3.27×10^{-3}	2.25×10^{-2}
cSARS Bat SRBD 60Hour GSE37827	8/300	3.27×10^{-3}	2.20×10^{-2}
icSARS CoV 0Hour GSE37827	8/300	3.27×10^{-3}	2.16×10^{-2}
icSARS CoV 48Hour GSE37827	8/300	3.27×10^{-3}	2.11×10^{-2}

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 April 2020 doi:10.20944/preprints202004.0524.v1

Table 5
Genes whose expression was altered by SARS-CoV-2-related viruses that significantly interact with the 163 genes selected by TD-based unsupervised FE because of "Virus Perturbations from GEO down" in Enrichr

Term	Overlap	P-value	Adjusted P-value
SARS-CoV 0Hour GSE47961	14/300	1.76×10^{-7}	1.42×10^{-5}
SARS-ddORF6 0Hour GSE47961	10/300	1.82×10^{-4}	4.51×10^{-3}
SARS-BatSRBD 96Hour GSE47960	9/300	8.14×10^{-4}	1.38×10^{-2}
SARS-CoV 24Hour GSE17400	9/300	8.14×10^{-4}	1.31×10^{-2}
cSARS Bat SRBD 60Hour GSE37827	9/300	8.14×10^{-4}	1.25×10^{-2}
icSARS CoV 48Hour GSE37827	9/300	8.14×10^{-4}	1.19×10^{-2}
SARS-CoV MA15 Day4-PFU-10^2 GSE33266	8/300	3.27×10^{-3}	3.11×10^{-2}

Alvocidib Alvocidib showed the sixth smallest (significant) *P*-value in "LINCS L1000 Chem Pert down" and had multiple hits (Tables S15 and S16). Alvocidib, a kinase inhibitor, was repurposing as an antiviral agent to control influenza A virus replication [17].

Mitoxantrone Mitoxantrone exhibited the 20th smallest (significant) *P*-value in "LINCS L1000 Chem Pert down" and had multiple hits (Tables S17 and S18). Mitoxantrone inhibits Vaccinia virus replication by blocking virion assembly [8].

QL-XII-47 QL-XII-47 showed the 22nd smallest (significant) *P*-value in "LINCS L1000 Chem Pert down" and had multiple hits (Tables S19 and S20). QL-XII-47's inhibition of Zika virus, West Nile virus, hepatitis C virus, and poliovirus have been reported previously [28].

Geldanamycin Geldanamycin showed the 25th smallest (significant) *P*-value in "LINCS L1000 Chem Pert down" and had multiple hits (Tables S21 and S22). Similar to radicicol as described above, the antiviral activity and RNA polymerase of radicicol involves degradation following Hsp90 inhibition in a range of negative-strand viruses [6]. These observations for radicicol are also applicable to geldanamycin.

3.3.2. *Drug perturbations from GEO*

Although we successfully identified numerous drug candidate compounds, it would also be useful to identify more candidates in other categories to confirm the effectiveness of our strategy. Thus, we next investigate "Drug Perturbations from GEO up/down" categories. As described below, we found numerous drug candidate compounds within these data sets (Table 6).

Fluticasone Effect of fluticasone propionate on virus-induced airway inflammation and antiviral immune responses in mice [19].

Atorvastatin Atorvastatin restricts the ability of influenza virus to generate lipid droplets and severely suppresses virus replication [9].

Quercetin Quercetin was reported to inhibit the cell entry of SARS-CoV-2 [32] and was included in the list of candidate compounds for SARS-CoV-2 screened by an *in silico* method [27].

Motexafin gadolinium Motexafin gadolinium was reported to selectively induce apoptosis in HIV-1-infected CD4+ T helper cells [16].

Trovafloxacin Simian virus 40 large T antigen helicase activity was inhibited by fluoroquinolone, trovafloxacin [1].

Doxycycline Antiviral activity of doxycycline against vesicular stomatitis virus was observed *in vitro* [30].

3.3.3. *Drug matrix*

To further confirm the independency of our findings based on the data sets used, we also examined the "Drug Matrix" category (Table 7, the full list is available in the supplementary materials). As we found some hits, our method can robustly identify promising drug candidate compounds.

Meloxicam Meloxicam is known to exert cytotoxic and antiproliferative activities towards virus-transformed tumor cells [7], including myelocytomatosis virus and Rous sarcoma virus. Myelocytomatosis virus is a retrovirus, which is an enveloped, negative-sense, single-stranded RNA virus, whereas Rous sarcoma virus is an enveloped, positive-sense, single-stranded RNA virus.

Gentamicin Although gentamicin is known to be a bactericidal antibiotic, it also exhibits antiviral activity (Table 3 [10]).

Dibromochloromethane Dibromochloromethane was announced as a possible antiviral drug by the Agency for Toxic Substances and Disease Registry (PUBLIC HEALTH STATEMENT Bromoform and Dibromochloromethane CAS#: 75-25-2 and 124-48-1, 2005)

3.4. Comparison with *in silico* drug discovery

Finally, we compared our results with those of other drugs identified *in silico*. As expected, some overlap was observed.

3.4.1. *Comparison with Wu et al. [29]*

We found multiple hits, which are summarized in Table 8; Wu et al. [29] identified 29 potential PLpro inhibitors,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 April 2020 doi:10.20944/preprints202004.0524.v1

Table 6
Genes whose expression is altered by SARS-CoV-2-related viruses that significantly interact with the 163 genes selected by TD-based unsupervised FE due to "Drug Perturbations from GEO up/down" in Enrichr

Term	Overlap	P-value	Adjusted P-value
Drug Perturbations from GEO up			
MK-886 CID 3651377 human GSE3202 sample 3193	54/368	3.90×10^{-53}	3.53×10^{-50}
fluticasone 5311101 human GSE15823 sample 3090	53/351	8.70×10^{-53}	3.94×10^{-50}
1-Naphthyl isothiocyanate 11080 rat GSE5509 sample 3568	50/301	7.71×10^{-52}	2.33×10^{-49}
quercetin 5280343 human GSE7259 sample 3416	50/327	6.03×10^{-50}	1.36×10^{-47}
N-METHYLFORMAMIDE 31254 rat GSE5509 sample 3570	46/283	4.37×10^{-47}	7.93×10^{-45}
NICKEL 935 human GSE6907 sample 3531	46/288	1.02×10^{-46}	1.54×10^{-44}
apratoxin A 6326668 human GSE2742 sample 3070	43/246	2.30×10^{-45}	2.98×10^{-43}
quercetin 5280343 human GSE7259 sample 3415	47/336	6.05×10^{-45}	6.85×10^{-43}
sapphyrin PCI-2050 (1.25 Î¼M) 9855235 human GSE6400 sample 3101	48/367	1.71×10^{-44}	1.72×10^{-42}
rosiglitazone DB00412 human GSE7035 sample 2810	43/281	9.52×10^{-43}	8.63×10^{-41}
Drug Perturbations from GEO down			
gatifloxacin 5379 human GSE9166 sample 2626	48/266	1.61×10^{-51}	1.46×10^{-48}
atorvastatin DB01076 human GSE2450 sample 2484	46/250	1.02×10^{-49}	4.62×10^{-47}
bexarotene DB00307 human GSE12791 sample 2681	46/253	1.84×10^{-49}	5.54×10^{-47}
clinafloxacin 60063 human GSE9166 sample 2625	55/470	1.30×10^{-48}	2.95×10^{-46}
motexafin gadolinium (12 h) DB05428 human GSE2189 sample 3127	48/320	1.89×10^{-47}	3.41×10^{-45}
BPDE 41322 human GSE19510 sample 3379	47/300	2.36×10^{-47}	3.55×10^{-45}
trovafloxacin 62959 human GSE9166 sample 2629	53/459	1.98×10^{-46}	2.55×10^{-44}
HYPOCHLOROUS ACID 24341 human GSE11630 sample 3199	40/204	2.85×10^{-44}	3.21×10^{-42}
trovafloxacin DB00685 human GSE9166 sample 3036	51/451	4.05×10^{-44}	4.06×10^{-42}
doxycycline DB00254 human GSE2624 sample 3077	48/391	3.82×10^{-43}	3.45×10^{-41}

Table 7
Genes whose expression is altered by SARS-CoV-2-related viruses that significantly interact with the 163 genes selected by TD-based unsupervised FE with "Drug Matrix" in Enrichr

Term	Overlap	P-value	Adjusted P-value
2-Amino-4-Nitrophenol-625 mg/kg in CMC-Rat-Kidney-1d-up	26/300	2.01×10^{-19}	1.59×10^{-15}
Allyl Alcohol-32 mg/kg in Saline-Rat-Liver-1d-up	25/291	1.30×10^{-18}	5.12×10^{-15}
Meloxicam-33 mg/kg in Corn Oil-Rat-Kidney-5d-up	23/261	1.96×10^{-17}	5.14×10^{-14}
Lipopolysaccharide E. Coli O55:B5-1.25 mg/kg in Saline-Rat-Kidney-1d-up	24/295	2.36×10^{-17}	4.64×10^{-14}
44'-Methylenedianiline-81 mg/kg in Corn Oil-Rat-Liver-3d-up	25/333	3.27×10^{-17}	5.16×10^{-14}
Gentamicin-40 mg/kg in Saline-Rat-Kidney-14d-up	24/309	6.83×10^{-17}	8.96×10^{-14}
Lead(IV) Acetate-600 mg/kg in Saline-Rat-Kidney-5d-up	24/309	6.83×10^{-17}	7.68×10^{-14}
Dibromochloromethane-325 mg/kg in CMC-Rat-Kidney-3d-up	24/312	8.51×10^{-17}	8.38×10^{-14}
Allopurinol-175 mg/kg in Corn Oil-Rat-Kidney-3d-up	24/329	2.84×10^{-16}	2.49×10^{-13}
Benzyl Acetate-1868 mg/kg in CMC-Rat-Kidney-3d-up	24/330	3.05×10^{-16}	2.40×10^{-13}

27 potential 3CLpro inhibitors, and 20 potential RdRp inhibitors from the ZINC drug database, and identified 13 potential PLpro inhibitors, 26 potential 3Clpro inhibitors, and 20 Potential RdRp inhibitors from their in-house natural product database. Doxycycline was among both the potential PLpro and 3CLpro inhibitors; ascorbic acid and isotretinoin were among the potential PLpro inhibitors; pioglitazone was among the potential 3CLpro inhibitors; and cortisone and tibolone were included as potential RdRp inhibitors from the ZINC drug database. These multiple hits further support the suitability of our strategy.

3.4.2. Comparison with Ubani et al. [27]

Ubani et al. [27] screened a library of 22 phytochemicals with antiviral activity obtained from the PubChem database for activity against the spike envelope glycoprotein and main protease of SARS-CoV-2. Among these, we found only one hit that overlapped with our screened out drugs, which was quercetin (Table 9).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 April 2020 doi:10.20944/preprints202004.0524.v1

Table 8
List of *in silico* screened drugs [29] whose target genes were also enriched in the 163 genes selected by TD-based unsupervised FE.

Term	Overlap	P-value	Adjusted P-value
Drug Perturbations from GEO up			
doxycycline DB00254 human GSE2624 sample 3076	38/272	3.93×10^{-36}	1.32×10^{-34}
doxycycline DB00254 human GSE2624 sample 3075	28/242	2.49×10^{-24}	2.02×10^{-23}
doxycycline DB00254 human GSE2624 sample 3077	23/209	1.30×10^{-19}	6.43×10^{-19}
doxycycline DB00254 mouse GSE29848 sample 3208	25/291	1.30×10^{-18}	5.84×10^{-18}
doxycycline DB00254 mouse GSE29848 sample 3209	24/267	2.35×10^{-18}	1.03×10^{-17}
doxycycline DB00254 human GSE2624 sample 3074	16/175	1.07×10^{-12}	2.89×10^{-12}
doxycycline DB00254 mouse GSE29848 sample 3207	17/225	4.54×10^{-12}	1.16×10^{-11}
ascorbic acid 54670067 human GSE11919 sample 3190	15/313	4.42×10^{-8}	8.64×10^{-8}
isotretinoin DB00982 human GSE10432 sample 2772	19/308	8.45×10^{-12}	2.10×10^{-11}
isotretinoin 5282379 human GSE10433 sample 2498	10/245	3.39×10^{-5}	5.51×10^{-5}
pioglitazone DB01132 rat GSE21329 sample 2843	40/400	3.44×10^{-32}	7.08×10^{-31}
pioglitazone DB01132 rat GSE21329 sample 2842	20/349	8.84×10^{-12}	2.18×10^{-11}
pioglitazone 4829 mouse GSE1458 sample 2587	19/318	1.47×10^{-11}	3.55×10^{-11}
pioglitazone DB01132 rat GSE20219 sample 2794	18/292	3.13×10^{-11}	7.40×10^{-11}
pioglitazone DB01132 human GSE8157 sample 2796	13/331	3.36×10^{-6}	5.89×10^{-6}
pioglitazone DB01132 rat GSE21329 sample 2841	11/279	1.88×10^{-5}	3.11×10^{-5}
pioglitazone DB01132 rat GSE20219 sample 2795	9/330	1.58×10^{-3}	2.31×10^{-3}
Drug Perturbations from GEO down			
doxycycline DB00254 human GSE2624 sample 3077	48/391	3.82×10^{-43}	3.45×10^{-41}
doxycycline DB00254 human GSE2624 sample 3074	39/425	6.14×10^{-30}	9.09×10^{-29}
doxycycline DB00254 human GSE2624 sample 3076	30/328	5.30×10^{-23}	4.02×10^{-22}
doxycycline DB00254 human GSE2624 sample 3075	27/358	1.40×10^{-18}	6.83×10^{-18}
doxycycline DB00254 mouse GSE29848 sample 3207	21/375	3.98×10^{-12}	1.21×10^{-11}
doxycycline DB00254 mouse GSE29848 sample 3208	16/309	5.14×10^{-9}	1.21×10^{-8}
doxycycline DB00254 mouse GSE29848 sample 3209	14/333	6.21×10^{-7}	1.28×10^{-6}
ascorbic acid 54670067 human GSE11919 sample 3190	40/287	5.09×10^{-38}	1.84×10^{-36}
isotretinoin DB00982 human GSE10432 sample 2772	7/292	1.02×10^{-2}	1.57×10^{-2}
pioglitazone DB01132 rat GSE21329 sample 2841	43/321	3.57×10^{-40}	1.90×10^{-38}
pioglitazone 4829 mouse GSE1458 sample 2587	24/282	8.34×10^{-18}	3.77×10^{-17}
pioglitazone DB01132 rat GSE21329 sample 2842	18/251	2.50×10^{-12}	7.64×10^{-12}
pioglitazone DB01132 rat GSE20219 sample 2794	17/308	6.28×10^{-10}	1.62×10^{-9}
pioglitazone DB01132 human GSE8157 sample 2796	14/269	4.58×10^{-8}	1.02×10^{-7}
pioglitazone DB01132 rat GSE20219 sample 2795	12/270	2.29×10^{-6}	4.52×10^{-6}
pioglitazone DB01132 rat GSE21329 sample 2843	7/200	1.29×10^{-3}	2.14×10^{-3}
tibolone 444008 human GSE12446 sample 3204	30/313	1.34×10^{-23}	1.14×10^{-22}

4. Discussion and Conclusion

In this study, we propose an advanced unsupervised learning method working in 4D tensors for identifying numerous promising drug candidate compounds for treating COVID-19 infection. The proposed method works by applying TD-based unsupervised FE to gene expression profiles of multiple lung cancer cell lines infected by SARS-CoV-2. We successfully identified 163 human genes predicted to be involved in the SARS-CoV-2 infection process. By uploading these selected 163 genes to Enrichr, we found that numerous drug compounds significantly altered expression of the genes.

Various analyses demonstrated that our results are robust. First, in a previous study [25] in which we employed a similar strategy to understand the infectious process of mouse hepatitis virus, a well-studied model CoV, we also identifies numerous drug candidate compounds in "DrugMatrix" and "Drug Pert from GEO up/down" categories in Enrichr. Although these drug compounds identified in the previous study are not always identified as top-ranked categories in this study (Tables 6 and 7), most were also significant. For example, in the "Drug Matrix" category, the identified drugs in the previous study were primaquine, meloxicam, cytarabine, pyrogallol, catechol, and neomycin. Among these six drugs, none, except for meloxicam, were ranked within the top ten (Table 7) but still significantly affected the expression of the selected 163 genes in this study (Table 10).

In the "Drug Pert from GEO up/down" category, the identified drugs in the previous study were fenretinide, pioglitazone, quercetin, decitabine, troglitazone, and motexafin gadolinium. Among these, only quercetin and motexafin gadolinium were identified in the present study (Table 6) and significantly affected the expression of the selected 163 genes (Table 11).

Additionally, doxycycline, ascorbic acid, isotretinoin, pi-

Table 9
List of *in silico* screened drugs [27] whose target genes are also enriched in the 163 genes selected by TD based unsupervised FE.

Term	Overlap	P-value	Adjusted P-value
Drug Perturbations from GEO up			
quercetin 5280343 human GSE7259 sample 3416	50/327	6.03×10^{-50}	1.36×10^{-47}
quercetin 5280343 human GSE7259 sample 3415	47/336	6.05×10^{-45}	6.85×10^{-43}
quercetin 5280343 rat GSE7479 sample 3409	38/394	5.73×10^{-30}	9.80×10^{-29}
quercetin 5280343 human GSE13899 sample 3182	19/307	7.99×10^{-12}	1.99×10^{-11}
quercetin DB04216 mouse GSE38136 sample 3436	17/297	3.59×10^{-10}	7.85×10^{-10}
quercetin DB04216 mouse GSE38141 sample 3435	16/280	1.25×10^{-9}	2.67×10^{-9}
quercetin DB04216 mouse GSE38136 sample 3438	15/254	2.69×10^{-9}	5.66×10^{-9}
quercetin DB04216 mouse GSE38067 sample 3440	13/227	4.62×10^{-8}	9.01×10^{-8}
quercetin DB04216 mouse GSE38136 sample 3437	16/472	1.66×10^{-6}	2.96×10^{-6}
quercetin DB04216 mouse GSE38067 sample 3441	7/114	4.16×10^{-5}	6.73×10^{-5}
quercetin DB04216 mouse GSE4262 sample 3428	11/360	1.85×10^{-4}	2.86×10^{-4}
quercetin DB04216 mouse GSE4262 sample 3429	8/229	5.94×10^{-4}	8.90×10^{-4}
quercetin DB04216 mouse GSE4262 sample 3427	9/360	2.84×10^{-3}	4.06×10^{-3}
quercetin DB04216 mouse GSE4262 sample 3433	8/323	5.09×10^{-3}	7.12×10^{-3}
quercetin DB04216 human GSE15162 sample 3444	7/323	1.69×10^{-2}	2.25×10^{-2}
quercetin DB04216 mouse GSE4262 sample 3434	7/324	1.71×10^{-2}	2.27×10^{-2}
Drug Perturbations from GEO down			
quercetin DB04216 mouse GSE38067 sample 3441	35/486	2.68×10^{-23}	2.11×10^{-22}
quercetin 5280343 human GSE13899 sample 3182	28/293	5.05×10^{-22}	3.40×10^{-21}
quercetin 5280343 rat GSE7479 sample 3409	16/206	1.31×10^{-11}	3.90×10^{-11}
quercetin DB04216 mouse GSE38141 sample 3435	17/320	1.13×10^{-9}	2.79×10^{-9}
quercetin DB04216 mouse GSE38136 sample 3436	16/303	3.89×10^{-9}	9.26×10^{-9}
quercetin DB04216 mouse GSE38067 sample 3440	15/373	4.27×10^{-7}	8.83×10^{-7}
quercetin 5280343 human GSE7259 sample 3415	12/264	1.81×10^{-6}	3.59×10^{-6}
quercetin DB04216 mouse GSE38136 sample 3438	13/346	5.44×10^{-6}	1.05×10^{-5}
quercetin DB04216 mouse GSE38136 sample 3437	8/128	1.02×10^{-5}	1.92×10^{-5}
quercetin DB04216 mouse GSE4262 sample 3430	11/312	5.22×10^{-5}	9.45×10^{-5}
quercetin DB04216 mouse GSE4262 sample 3431	10/348	5.87×10^{-4}	9.96×10^{-4}
quercetin 5280343 human GSE7259 sample 3416	8/273	1.83×10^{-3}	3.02×10^{-3}
quercetin DB04216 mouse GSE4262 sample 3428	7/240	3.59×10^{-3}	5.74×10^{-3}
quercetin DB04216 mouse GSE4262 sample 3429	7/371	3.27×10^{-2}	4.73×10^{-2}

oglitazone, cortisone, andtibolone, and quercetin were identified in the comparison with two other *in slico* studies. These drugs were also identified in the comparison between the present study and other *in slico* studies (Tables 8 and 9). The overlapping results with the previous study suggest that our strategy is quite robust.

These results are also thought to be biologically sound. For example, Although A-443654 is inhibitor of Akt, which is important for SARS-CoV infection (see above). Radicicol and geldanamycin inhibit Hsp90. The importance of inhibition of Hsp90 was reported for treating patients with COVID-19 has been reported previously [21]. Although we could not identify all biological meanings of the identified drugs, these two examples suggest that the results are biologically sound.

CRediT authorship contribution statement

Y-h. Taguchi: Conceptualization of this study, Methodology, Software, Writing - Original draft preparation. **Turki Turki:** Conceptualization of this study, Data preparation, Writing - Original draft preparation.

References

[1] Ali, S.H., Chandraker, A., DeCaprio, J.A., 2007. Inhibition of simian virus 40 large t antigen. Antivir Ther 12, 1–6.

[2] Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.C., Møller, R., Panis, M., Sachs, D., Albrecht, R.A., tenOever, B.R., 2020. Sars-cov-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. bioRxiv URL: https://www.biorxiv.org/content/early/2020/03/24/2020.03.24.004655, doi:10.1101/2020.03.24.004655, arXiv:https://www.biorxiv.org/content/early/2020/03/24/2020.03.24.004655.full.pd

[3] Blázquez, A.B., Vázquez-Calvo, A., Martin-Acebes, M.A., Saiz, J.C., 2018. Pharmacological inhibition of protein kinase C reduces West Nile virus replication. Viruses 10. URL: https://www.mdpi.com/1999-4915/10/2/91, doi:10.3390/v10020091.

[4] Brahms, A., Mudhasani, R., Pinkham, C., Kota, K., Nasar, F., Zamani, R., Bavari, S., Kehn-Hall, K., 2017. Sorafenib impedes rift valley fever virus egress by inhibiting valosin-containing protein function in the cellular secretory pathway. Journal of Virology 91. URL: https://jvi.asm.org/content/91/21/e00968-17, doi:10.1128/JVI.00968-17, arXiv:https://jvi.asm.org/content/91/21/e00968-17.full.pdf.

[5] Chan, C.M., Ma, C.W., Chan, W.Y., Chan, H.Y.E., 2007. The sars-coronavirus membrane protein induces apoptosis through modulating the akt survival pathway. Archives of Biochemistry and Biophysics 459, 197 – 207. URL: http://www.sciencedirect.com/science/article/pii/S000398610700015X, doi:https://doi.org/10.

Table 10
Five Drugs ranked within top 10 in the previous study but not in the present study in "DrugMatrix" category in Enrichr. They were still significantly enriched for the selected 163 genes. If there were more than ten hits, they were omitted.

Term	Overlap	P-value	Adjusted P-value
Primaquine-45 mg/kg in CMC-Rat-Liver-5d-up	18/315	1.09×10^{-10}	3.34×10^{-9}
Primaquine-45 mg/kg in CMC-Rat-Liver-1d-up	15/337	1.16×10^{-7}	7.37×10^{-7}
Primaquine-45 mg/kg in CMC-Rat-Liver-3d-up	14/316	3.31×10^{-7}	1.66×10^{-6}
Primaquine-45 mg/kg in CMC-Rat-Liver-3d-dn	9/284	5.51×10^{-4}	8.62×10^{-4}
Primaquine-45 mg/kg in CMC-Rat-Liver-5d-dn	7/285	8.98×10^{-3}	1.06×10^{-2}
Cytarabine-487 mg/kg in Saline-Rat-Bone marrow-1d-up	17/326	1.49×10^{-9}	2.47×10^{-8}
Cytarabine-23 mg/kg in Saline-Rat-Liver-0.25d-up	16/313	6.17×10^{-9}	7.17×10^{-8}
Cytarabine-487 mg/kg in Saline-Rat-Liver-1d-dn	14/237	9.28×10^{-9}	9.81×10^{-8}
Cytarabine-23 mg/kg in Saline-Rat-Bone marrow-0.25d-up	17/385	1.79×10^{-8}	1.66×10^{-7}
Cytarabine-23 mg/kg in Saline-Rat-Spleen-3d-up	15/299	2.42×10^{-8}	2.12×10^{-7}
Cytarabine-487 mg/kg in Saline-Rat-Liver-5d-dn	14/291	1.21×10^{-7}	7.59×10^{-7}
Cytarabine-23 mg/kg in Saline-Rat-Liver-5d-dn	14/307	2.33×10^{-7}	1.26×10^{-6}
Cytarabine-487 mg/kg in Saline-Rat-Kidney-5d-dn	14/319	3.71×10^{-7}	1.84×10^{-6}
Cytarabine-487 mg/kg in Saline-Rat-Kidney-3d-dn	14/327	4.99×10^{-7}	2.35×10^{-6}
Cytarabine-487 mg/kg in Saline-Rat-Spleen-1d-up	14/329	5.37×10^{-7}	2.49×10^{-6}
Cytarabine-23 mg/kg in Saline-Rat-Spleen-0.25d-up	14/344	9.14×10^{-7}	3.83×10^{-6}
(additional 31 hits with less significance are omitted)			
Pyrogallol-1000 mg/kg in Water-Rat-Liver-5d-up	14/304	2.07×10^{-7}	1.14×10^{-6}
Pyrogallol-1000 mg/kg in Water-Rat-Liver-1d-up	15/409	1.35×10^{-6}	5.23×10^{-6}
Pyrogallol-1000 mg/kg in Water-Rat-Liver-5d-dn	12/296	5.88×10^{-6}	1.76×10^{-5}
Pyrogallol-1000 mg/kg in Water-Rat-Liver-3d-up	13/349	5.97×10^{-6}	1.78×10^{-5}
Pyrogallol-1000 mg/kg in Water-Rat-Liver-3d-dn	7/251	4.59×10^{-3}	5.69×10^{-3}
Pyrogallol-1000 mg/kg in Water-Rat-Liver-1d-dn	5/191	2.03×10^{-2}	2.26×10^{-2}
Catechol-195 mg/kg in Saline-Rat-Liver-0.25d-up	19/290	2.94×10^{-12}	2.41×10^{-10}
Catechol-40 mg/kg in Saline-Rat-Liver-0.25d-up	19/305	7.13×10^{-12}	4.16×10^{-10}
Catechol-195 mg/kg in Saline-Rat-Bone marrow-1d-dn	16/305	4.27×10^{-9}	5.49×10^{-8}
Catechol-40 mg/kg in Saline-Rat-Kidney-0.25d-dn	16/319	8.08×10^{-9}	8.82×10^{-8}
Catechol-40 mg/kg in Saline-Rat-Kidney-3d-dn	15/294	1.93×10^{-8}	1.77×10^{-7}
Catechol-40 mg/kg in Saline-Rat-Bone marrow-0.25d-dn	15/306	3.28×10^{-8}	2.67×10^{-7}
Catechol-40 mg/kg in Saline-Rat-Bone marrow-1d-dn	15/307	3.43×10^{-8}	2.76×10^{-7}
Catechol-195 mg/kg in Saline-Rat-Spleen-1d-up	15/320	5.91×10^{-8}	4.31×10^{-7}
Catechol-195 mg/kg in Saline-Rat-Kidney-5d-dn	14/281	7.87×10^{-8}	5.39×10^{-7}
Catechol-195 mg/kg in Saline-Rat-Bone marrow-5d-dn	14/310	2.62×10^{-7}	1.38×10^{-6}
(additional 27 hits with less significance are omitted)			
Neomycin-877 mg/kg in Corn Oil-Rat-Kidney-1d-dn	14/264	3.62×10^{-8}	2.88×10^{-7}
Neomycin-877 mg/kg in Corn Oil-Rat-Liver-5d-up	14/323	4.31×10^{-7}	2.08×10^{-6}
Neomycin-56 mg/kg in Corn Oil-Rat-Kidney-0.25d-dn	12/256	1.31×10^{-6}	5.12×10^{-6}
Neomycin-877 mg/kg in Corn Oil-Rat-Kidney-3d-up	13/311	1.69×10^{-6}	6.23×10^{-6}
Neomycin-56 mg/kg in Corn Oil-Rat-Kidney-5d-dn	12/270	2.29×10^{-6}	7.99×10^{-6}
Neomycin-56 mg/kg in Corn Oil-Rat-Liver-5d-up	12/279	3.21×10^{-6}	1.06×10^{-5}
Neomycin-56 mg/kg in Corn Oil-Rat-Kidney-3d-dn	11/233	3.43×10^{-6}	1.12×10^{-5}
Neomycin-877 mg/kg in Corn Oil-Rat-Kidney-3d-dn	12/289	4.60×10^{-6}	1.44×10^{-5}
Neomycin-877 mg/kg in Corn Oil-Rat-Liver-0.25d-dn	12/296	5.88×10^{-6}	1.76×10^{-5}
Neomycin-56 mg/kg in Corn Oil-Rat-Liver-3d-dn	12/309	9.07×10^{-6}	2.50×10^{-5}
(additional 20 hits with less significance are omitted)			

1016/j.abb.2007.01.012.

[6] Connor, J.H., McKenzie, M.O., Parks, G.D., Lyles, D.S., 2007. Antiviral activity and rna polymerase degradation following hsp90 inhibition in a range of negative strand viruses. Virology 362, 109 – 119. URL: http://www.sciencedirect.com/science/article/pii/S0042682206009330, doi:https://doi.org/10.1016/j.virol.2006.12.026.

[7] CULITA1, D.C., ALEXANDROVA, R., DYAKOVA, L., MARINESCU, G., PATRON, L., KALFIN, R., ALEXANDROV, M., 2012. Evaluation of cytotoxic and antiproliferative activity of co(ii), ni(ii), cu(ii) and zn(ii) complexes with meloxicam on virus – transformed tumor cells daniela. Revista de Chimie 63, 384–389.

[8] Deng, L., Dai, P., Ciro, A., Smee, D.F., Djaballah, H., Shuman, S., 2007. Identification of novel antipoxviral agents: Mitoxantrone inhibits vaccinia virus replication by blocking virion assembly. Journal of Virology 81, 13392–13402. URL: https://jvi.asm.org/content/81/24/13392, doi:10.1128/JVI.00770-07, arXiv:https://jvi.asm.org/content/81/24/13392.full.pdf.

[9] Episcopio, D., Aminov, S., Benjamin, S., Germain, G., Datan, E., Landazuri, J., Lockshin, R.A., Zakeri, Z., 2019. Atorvastatin restricts

Table 11
Four Drugs ranked within top 10 in the previous study but not in the present study in "Drug Pert from GEO up/down: category in Enrichr. They were still significantly enriched toward the selected 163 genes.

Term	Overlap	P-value	Adjusted P-value
Drug Perturbations from GEO up			
fenretinide 5288209 rat GSE3952 sample 3561	35/397	2.98×10^{-26}	2.81×10^{-25}
fenretinide 5288209 rat GSE3952 sample 3559	7/160	3.45×10^{-4}	5.30×10^{-4}
pioglitazone DB01132 rat GSE21329 sample 2843	40/400	3.44×10^{-32}	7.08×10^{-31}
pioglitazone DB01132 rat GSE21329 sample 2842	20/349	8.84×10^{-12}	2.18×10^{-11}
pioglitazone 4829 mouse GSE1458 sample 2587	19/318	1.47×10^{-11}	3.55×10^{-11}
pioglitazone DB01132 rat GSE20219 sample 2794	18/292	3.13×10^{-11}	7.40×10^{-11}
pioglitazone DB01132 human GSE8157 sample 2796	13/331	3.36×10^{-6}	5.89×10^{-6}
pioglitazone DB01132 rat GSE21329 sample 2841	11/279	1.88×10^{-5}	3.11×10^{-5}
pioglitazone DB01132 rat GSE20219 sample 2795	9/330	1.58×10^{-3}	2.31×10^{-3}
decitabine DB01262 human GSE29077 sample 2546	31/243	3.22×10^{-28}	3.99×10^{-27}
decitabine DB01262 human GSE29077 sample 2538	31/263	3.84×10^{-27}	4.05×10^{-26}
decitabine DB01262 human GSE9118 sample 2703	31/271	9.77×10^{-27}	9.73×10^{-26}
decitabine DB01262 human GSE29077 sample 2539	26/279	3.22×10^{-20}	1.69×10^{-19}
decitabine 451668 mouse GSE4768 sample 3103	25/251	3.55×10^{-20}	1.85×10^{-19}
decitabine 451668 mouse GSE4768 sample 3105	26/287	6.59×10^{-20}	3.36×10^{-19}
decitabine DB01262 human GSE29077 sample 2540	19/300	5.34×10^{-12}	1.35×10^{-11}
decitabine DB01262 human GSE29077 sample 2547	19/304	6.73×10^{-12}	1.69×10^{-11}
decitabine DB01262 human GSE29077 sample 2548	19/316	1.32×10^{-11}	3.21×10^{-11}
decitabine 451668 mouse GSE4768 sample 3108	12/374	5.91×10^{-5}	9.43×10^{-5}
troglitazone DB00197 rat GSE21329 sample 2833	36/408	5.13×10^{-27}	5.34×10^{-26}
troglitazone DB00197 rat GSE21329 sample 2834	28/198	8.28×10^{-27}	8.42×10^{-26}
troglitazone 5591 mouse GSE1458 sample 2589	26/305	3.05×10^{-19}	1.45×10^{-18}
troglitazone DB00197 rat GSE21329 sample 2832	10/245	3.39×10^{-5}	5.52×10^{-5}
Drug Perturbations from GEO down			
fenretinide 5288209 rat GSE3952 sample 3559	38/440	3.49×10^{-28}	4.56×10^{-27}
fenretinide 5288209 rat GSE3952 sample 3561	22/203	1.18×10^{-18}	5.84×10^{-18}
pioglitazone DB01132 rat GSE21329 sample 2841	43/321	3.57×10^{-40}	1.90×10^{-38}
pioglitazone 4829 mouse GSE1458 sample 2587	24/282	8.34×10^{-18}	3.77×10^{-17}
pioglitazone DB01132 rat GSE21329 sample 2842	18/251	2.50×10^{-12}	7.64×10^{-12}
pioglitazone DB01132 rat GSE20219 sample 2794	17/308	6.28×10^{-10}	1.62×10^{-9}
pioglitazone DB01132 human GSE8157 sample 2796	14/269	4.58×10^{-8}	1.02×10^{-7}
pioglitazone DB01132 rat GSE20219 sample 2795	12/270	2.29×10^{-6}	4.52×10^{-6}
pioglitazone DB01132 rat GSE21329 sample 2843	7/200	1.29×10^{-3}	2.14×10^{-3}
decitabine DB01262 human GSE29077 sample 2540	44/300	6.35×10^{-43}	5.21×10^{-41}
decitabine DB01262 human GSE29077 sample 2539	41/321	2.15×10^{-37}	7.19×10^{-36}
decitabine 451668 mouse GSE4768 sample 3108	35/226	6.98×10^{-35}	1.91×10^{-33}
decitabine DB01262 human GSE29077 sample 2538	37/337	3.06×10^{-31}	5.22×10^{-30}
decitabine DB01262 human GSE9118 sample 2703	29/329	8.47×10^{-22}	5.54×10^{-21}
decitabine DB01262 human GSE29077 sample 2548	25/284	7.22×10^{-19}	3.62×10^{-18}
decitabine DB01262 human GSE29077 sample 2547	21/296	4.08×10^{-14}	1.42×10^{-13}
decitabine 451668 mouse GSE4768 sample 3105	20/313	1.20×10^{-12}	3.76×10^{-12}
decitabine 451668 mouse GSE4768 sample 3103	16/349	2.85×10^{-8}	6.43×10^{-8}
decitabine DB01262 human GSE29077 sample 2546	14/357	1.42×10^{-6}	2.85×10^{-6}
troglitazone DB00197 rat GSE21329 sample 2832	37/355	2.09×10^{-30}	3.31×10^{-29}
troglitazone 5591 mouse GSE1458 sample 2589	17/295	3.24×10^{-10}	8.55×10^{-10}
troglitazone DB00197 rat GSE21329 sample 2834	16/402	1.98×10^{-7}	4.20×10^{-7}
troglitazone DB00197 rat GSE21329 sample 2833	11/192	5.16×10^{-7}	1.07×10^{-6}

the ability of influenza virus to generate lipid droplets and severely suppresses the replication of the virus. The FASEB Journal 33, 9516–9525. URL: https://doi.org/10.1096/fj.201900428RR, doi:10.1096/fj.201900428RR, arXiv:https://doi.org/10.1096/fj.201900428RR. pMID: 31125254.

[10] Fischer, A.B., 1975. Gentamicin as a bactericidal antibiotic in tissue culture. Medical Microbiology and Immunology 161, 23–39. URL: https://doi.org/10.1007%2Fbf02120767, doi:10.1007/bf02120767.

[11] He, B., Garmire, L., 2020. Repurposed drugs for treating lung injury in covid-19. arXiv:arXiv:2003.14333.

[12] Jeong, G., Ahn, B.Y., 2019. Aurora kinase a promotes hepatitis b virus replication and expression. Antiviral Research 170. doi:10.1016/j.antiviral.2019.104572.

[13] Kubota, N., 2016. Hepatitis c virus inhibitor synergism suggests

multistep interactions between heat-shock protein 90 and hepatitis c virus replication. World Journal of Hepatology 8, 282. URL: https://doi.org/10.4254%2Fwjh.v8.i5.282, doi:10.4254/wjh.v8.i5.282.

[14] 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., Ma'ayan, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research 44, W90–W97. URL: https://doi.org/10.1093/nar/gkw377, doi:10.1093/nar/gkw377, arXiv:https://academic.oup.com/nar/article-pdf/44/W1/W90/18788036/gkw377.pdf.

[15] Li, Y.H., Tao, P.Z., Liu, Y.Z., Jiang, J.D., 2004. Geldanamycin, a ligand of heat shock protein 90, inhibits the replication of herpes simplex virus type 1 in vitro. Antimicrobial Agents and Chemotherapy 48, 867–872. URL: https://aac.asm.org/content/48/3/867, doi:10.1128/AAC.48.3.867-872.2004, arXiv:https://aac.asm.org/content/48/3/867.full.pdf.

[16] Perez, O.D., Nolan, G.P., Magda, D., Miller, R.A., Herzenberg, L.A., Herzenberg, L.A., 2002. Motexafin gadolinium (gd-tex) selectively induces apoptosis in hiv-1 infected cd4+ t helper cells. Proceedings of the National Academy of Sciences 99, 2270–2274. URL: https://www.pnas.org/content/99/4/2270, doi:10.1073/pnas.261711499, arXiv:https://www.pnas.org/content/99/4/2270.full.pdf.

[17] Perwitasari, O., Yan, X., O'Donnell, J., Johnson, S., Tripp, R.A., 2015. Repurposing kinase inhibitors as antiviral agents to control influenza a virus replication. ASSAY and Drug Development Technologies 13, 638–649. URL: https://doi.org/10.1089/adt.2015.0003.drrr, doi:10.1089/adt.2015.0003.drrr, arXiv:https://doi.org/10.1089/adt.2015.0003.drrr. pMID: 26192013.

[18] Robson, B., 2020. Computers and viral diseases. preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the sars-cov-2 (2019-ncov, covid-19) coronavirus. Computers in Biology and Medicine , 103670.

[19] Singanayagam, A., Glanville, N., Bartlett, N., Johnston, S., 2015. Effect of fluticasone propionate on virus-induced airways inflammation and anti-viral immune responses in mice. The Lancet 385, S88. URL: https://doi.org/10.1016%2Fs0140-6736%2815%2960403-2, doi:10.1016/s0140-6736(15)60403-2.

[20] rong Su, A., Qiu, M., lei Li, Y., tao Xu, W., wei Song, S., hui Wang, X., yong Song, H., Zheng, N., wei Wu, Z., 2017. BX-795 inhibits HSV-1 and HSV-2 replication by blocking the JNK/p38 pathways without interfering with PDK1 activity in host cells. Acta Pharmacologica Sinica 38, 402–414. URL: https://doi.org/10.1038%2Faps.2016.160, doi:10.1038/aps.2016.160.

[21] Sultan, I., Howard, S., Tbakhi, A., 2020. Drug repositioning suggests a role for the heat shock protein 90 inhibitor geldanamycin in treating covid-19 infection doi:0.21203/rs.3.rs-18714/v1.

[22] Taguchi, Y., 2019. Drug candidate identification based on gene expression of treated cells using tensor decomposition-based unsupervised feature extraction for large-scale data. BMC bioinformatics 19, 388.

[23] h. Taguchi, Y., 2020. Unsupervised feture extraction applied to bioinformatics: PCA and TD based approach. Springer International, Switzerland.

[24] Taguchi, Y., Turki, T., 2019. Neurological disorder drug discovery from gene expression with tensor decomposition. Current Pharmaceutical Design .

[25] Taguchi, Y., Turki, T., 2020. Novel method for detection of genes with altered expression caused by coronavirus infection and screening of candidate drugs for sars-cov-2. preprints.org , 2020040431doi:10.20944/preprints202004.0431.v1.

[26] Taguchi, Y.H., 2017. Identification of candidate drugs using tensor-decomposition-based unsupervised feature extraction in integrated analysis of gene expression between diseases and drugmatrix datasets. Scientific reports 7, 1–16.

[27] Ubani, A., Agwom, F., Shehu, N.Y., Luka, P., Umera, A., Umar, U., Omale, S., Nnadi, N.E., Aguiyi, J.C., 2020. Molecular docking analysis of some phytochemicals on two sars-cov-2 targets. bioRxiv URL: https://www.biorxiv.org/content/early/2020/04/01/2020.03.31.017657, doi:10.1101/2020.03.31.017657, arXiv:https://www.biorxiv.org/content/early/2020/04/01/2020.03.31.017657.full.pdf.

[28] [de Wispelaere], M., Carocci, M., Liang, Y., Liu, Q., Sun, E., Vetter, M.L., Wang, J., Gray, N.S., Yang, P.L., 2017. Discovery of host-targeted covalent inhibitors of dengue virus. Antiviral Research 139, 171 – 179. URL: http://www.sciencedirect.com/science/article/pii/S0166354216306490, doi:https://doi.org/10.1016/j.antiviral.2016.12.017.

[29] Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., Li, X., Zheng, M., Chen, L., Li, H., 2020. Analysis of therapeutic targets for sars-cov-2 and discovery of potential drugs by computational methods. Acta Pharmaceutica Sinica B URL: http://www.sciencedirect.com/science/article/pii/S2211383520302999, doi:https://doi.org/10.1016/j.apsb.2020.02.008.

[30] Wu, Z.c., Wang, X., Wei, J.c., Li, B.b., Shao, D.h., Li, Y.m., Liu, K., Shi, Y.y., Zhou, B., Qiu, Y.f., Ma, Z.y., 2015. Antiviral activity of doxycycline against vesicular stomatitis virus in vitro. FEMS Microbiology Letters 362. URL: https://doi.org/10.1093/femsle/fnv195, doi:10.1093/femsle/fnv195, arXiv:https://academic.oup.com/femsle/article-pdf/362/22/fnv195/23925204/fnv195.pdf. fnv195.

[31] Yang, H., Kim, S.K., Kim, M., Reche, P.A., Morehead, T.J., Damon, I.K., Welsh, R.M., Reinherz, E.L., 2005. Antiviral chemotherapy facilitates control of poxvirus infections through inhibition of cellular signal transduction. The Journal of Clinical Investigation 115, 379–387. URL: https://www.jci.org/articles/view/23220, doi:10.1172/JCI23220.

[32] Yi, L., Li, Z., Yuan, K., Qu, X., Chen, J., Wang, G., Zhang, H., Luo, H., Zhu, L., Jiang, P., Chen, L., Shen, Y., Luo, M., Zuo, G., Hu, J., Duan, D., Nie, Y., Shi, X., Wang, W., Han, Y., Li, T., Liu, Y., Ding, M., Deng, H., Xu, X., 2004. Small molecules blocking the entry of severe acute respiratory syndrome coronavirus into host cells. Journal of Virology 78, 11334–11339. URL: https://jvi.asm.org/content/78/20/11334, doi:10.1128/JVI.78.20.11334-11339.2004, arXiv:https://jvi.asm.org/content/78/20/11334.full.pdf.

[33] Zhao, D., Fukuyama, S., Sakai-Tagawa, Y., Takashita, E., Shoemaker, J.E., Kawaoka, Y., 2016. C646, a novel p300/CREB-binding protein-specific inhibitor of histone acetyltransferase, attenuates influenza A virus infection. Antimicrobial Agents and Chemotherapy 60, 1902–1906. URL: https://aac.asm.org/content/60/3/1902, doi:10.1128/AAC.02055-15, arXiv:https://aac.asm.org/content/60/3/1902.full.pdf.

Y-H. TAGUCHI received a B.S. degree in physics from the Tokyo Institute of Technology and a Ph.D. degree in physics from the Tokyo Institute of Technology. He is currently a full professor with the Department of Physics, Chuo University, Japan. His works have been published in leading journals such as Physical Review Letters, Bioinformatics, and Scientific Reports. His research interests include bioinformatics, machine learning, and nonlinear physics. He is also an editorial board member of PloS ONE, BMC Medical Genomics, Medicine (Lippincott Williams & Wilkins journal), BMC Research Notes, non-coding RNA (MDPI), and IPSJ Transaction on Bioinformatics.

TURKI TURKI received a B.S. in computer science from King Abdulaziz University, an M.S. in computer science from NYU.POLY, and a Ph.D. in computer science from the New Jersey Institute of Technology. He is currently an assistant professor with the Department of Computer Science, King Abdulaziz University, Saudi Arabia. His research interests include artificial intelli-

gence, machine learning, deep learning, data mining, data science, big data analytics, and bioinformatics. His research has been accepted and published in journals such as Frontiers in Genetics, BMC Genomics, BMC Systems Biology, Expert Systems with Applications, Computers in Biology and Medicine, and Current Pharmaceutical Design. He was awarded several distinction awards from the Deanship of Scientific Research at King Abdulaziz University. He is supported by King Abdulaziz University and is currently working on several biomedicine related projects. Dr. Turki has served on the program committees of several international conferences. Additionally, he is an editorial board member of Sustainable Computing: Informatics and Systems and Computers in Biology and Medicine.