Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A Scheme for Inferring Viral-Host Associations based on Codon Usage Patterns Identifies the Most Affected Signaling Pathways during COVID-19

Jayanta Kumar Das, Subhadip Chakraborty, Swarup Roy

PII: S1532-0464(21)00130-1
DOI: https://doi.org/10.1016/j.jbi.2021.103801
Reference: YJBIN 103801

To appear in: Journal of Biomedical Informatics

Accepted Date: 3 May 2021

Please cite this article as: Das, J.K., Chakraborty, S., Roy, S., A Scheme for Inferring Viral-Host Associations based on Codon Usage Patterns Identifies the Most Affected Signaling Pathways during COVID-19, Journal of Biomedical Informatics (2021), doi: https://doi.org/10.1016/j.jbi.2021.103801

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Elsevier Inc. All rights reserved.
A Scheme for Inferring Viral-Host Associations based on Codon Usage Patterns Identifies the Most Affected Signaling Pathways during COVID-19

Jayanta Kumar Das\textsuperscript{a}, Subhadip Chakraborty\textsuperscript{b}, Swarup Roy\textsuperscript{c,∗};

\textsuperscript{a}Department of Pediatrics, Johns Hopkins University School of Medicine, Maryland, USA
\textsuperscript{b}Department of Botany, Nabadvip Vidyasagar College, Nabadvip, India
\textsuperscript{c}Network Reconstruction & Analysis (NetRA) Lab, Department of Computer Applications, Sikkim University, Gangtok, India

Abstract

Understanding the molecular mechanism of COVID-19 disease pathogenesis helps in the rapid development of therapeutic targets. Usually, viral protein targets host proteins in an organized fashion. The expression of any viral gene depends mostly on the host translational machinery. Recent studies report the great significance of codon usage biases in establishing host-viral protein-protein interactions (PPI). Exploring the codon usage patterns between a pair of co-evolved host and viral proteins may present novel insight into the host-viral protein interactomes during disease pathogenesis. Leveraging the similarity (and dissimilarity) in codon usage patterns, we propose a computational scheme to recreate the host-viral protein-protein interaction network. We use host proteins from seventeen (17) essential signaling pathways for our current work towards understanding the possible targeting mechanism of SARS-CoV-2 viral proteins. We infer both negatively and positively interacting edges in the network. Further, extensive analysis is performed to understand the host PPI network topologically and the attacking behavior of the viral proteins. Our study reveals that viral proteins mostly utilize codons, rare in the targeted host proteins (negatively correlated interaction). Among non-structural proteins, NSP3 and structural protein, Spike (S) protein, are the most influential proteins in interacting multiple host proteins. While ranking the most affected pathways, MAPK pathways observe to be the worst affected during the COVID-19 disease. Several proteins participating in multiple pathways are highly central in host PPI and mostly targeted by multiple viral proteins. We observe many potential targets (host proteins) from the affected pathways associated with the various drugs molecules including Arsenic trioxide, Dexamethasone, Hydroxychloroquine, Ritonavir, and Interferon beta, which are either under trail or in use during COVID-19.

Keywords: Protein Interaction Network, Codon usage bias, Bipartite graph, Cell signaling, Relative Synonymous Codon Usage, Centrality, Drugs

∗Corresponding Author
Email address: sroy01@cus.ac.in (Swarup Roy)
1. Introduction

The entire world is passing through an unprecedented pandemic situation due to a massive outbreak of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infected viral disease, COVID-19. SARS-CoV-2 belongs to the Coronaviridae family, and members of this family are enveloped, non-segmented, single-stranded, positive-sense large RNA genomes [1]. This virus rapidly spreads from person to person by respiratory droplets during close physical contact. Other than respiratory organs, it is reported to attack the immune system, and other vital cellular machinery leads to multi-organ failure [2]. The need for the hour and utmost crucial for the scientific community to understand the disease pathogenesis of SARS-CoV-2 at genomic and proteomic levels for the rapid development of effective drugs or vaccines to control the COVID-19. Many recent works use host-viral protein-protein interaction network as an input to elucidate potential drug targets or repurposed drug molecules [3, 4, 5]. Host-pathogen protein interactions provide essential insights into the molecular mechanisms of pathogenicity [6].

The host defense mechanism activates signal transduction molecules that initiate signals, which activate immune effector mechanisms to protect the host from any pathogenic infections. Studies show that viral immune modulators perturb the human PPI network by targeting signaling pathways [7] to suppress the immunity in mammalian hosts [8]. To understand the molecular mechanism of pathogenicity of SARS-CoV-2 during COVID-19 disease, investigation of the host-viral protein interactions is important. Knowledge gained through understanding the interactions among viral and host proteins involved in signaling pathways may translate into effective therapies and vaccines. There are four (04) basic categories of chemical signaling found in multi-cellular organisms, namely, paracrine signaling, autocrine signaling, endocrine signaling, and signaling by direct contact. Various regulatory signaling pathways are involved in signaling transduction and cellular interactions, many of which are playing important role during COVID-19. Signal transduction focuses on molecular and functional aspects of viral interactions with host cell signaling, important for the anti-viral response, the viral life cycle, viral pathogenesis, and cell transformation [9]. We aim to study the interaction pattern of SARS-CoV-2 with its target host proteins involved in signaling pathways [10, 11, 12, 13, 14, 15] (see Materials and Method section). Working with them can help in deciphering the possible involvement of pathways and key genes during COVID-19 pathogenesis.

The virus-host interactome is essential for understanding virulence factors, influencing SARS-CoV-2 pathogenesis [16]. Recent studies reported the use of SARS-CoV-2 and host PPI networks to study the pathogenesis of SARS-CoV-2 and identify repurposed drugs [17, 4, 18]. Several studies have shown that viral proteins interact with hubs in complex host PPI network [19, 20]. Considering different properties of proteins such as sequence homology, gene co-expression, and phylogenetic profiles [21, 22, 23], the pairwise similarity is computed between a pair of proteins to predict a possible interaction between them. In addition to non-structural information, structural data about a pair of proteins appears to be more effective in improving prediction [24, 25, 4].
Several computational methods have been developed to predict PPIs by focusing on protein sequence features [26, 27, 28]. In reality, predicting whether two given proteins are physically interacting or not based on the similarity of different structural and non-structural features is challenging and not always feasible due to the expensive experimental setup. In case of virus genome study, codon usage biases play an important role [29, 30, 31]. Viral gene depends largely on the host transnational machinery for their expression [32]. Viral proteins are co-evolved with host proteins and several studies have reported that physically interacting or functionally associated protein pairs have similar codon usage bias [33, 34, 35, 36, 37, 38, 39]. Therefore, codon usage bias can be utilized in establishing host-viral protein interactions [40, 41]. State-of-the-art methods for inferring interactions do not consider the co-expression or co-adaptation between a pair of virus and its host proteins for drawing possible attacking mechanism of a virus [42, 43]. According to the genome hypothesis proposed by Grantham et al. [44], the pattern of codon usage is species-specific and in some way unique. Interestingly, even in the same genome, the codon usage varies significantly among genes with different expression levels [39], functions [45], and tissue-specific patterns [46]. Few works hypothesized that viral proteins enrich in codons, rare in their target host genomes [47, 48].

In this work, we explore host-viral interactions by leveraging the inherent correlation (co-expression or co-adaptation) between viral and host proteins’ codon usage patterns, which applies to any nucleotide (CDS) sequences. To the best of our knowledge, no prior work explored the codon usage similarity to infer host-viral PPI network. We capture both positive and negative interactions in the host-viral PPI. We use host proteins involved in different human cellular signaling pathways that might be having a role during COVID-19 disease pathogenesis. Topologically, we try to establish the relevance of the host proteins and highlight a few essential proteins in the network, which are also useful as drug targets for certain reported drugs during COVID-19.

2. Materials and Method

This section discusses proposed scheme for constructing a host-viral PPI network using pair-wise codon usage patterns of host and viral proteins. To analyze the interaction mechanism of SARS-CoV-2 viral proteins in host signaling pathways, we select all the genes involved in 17 candidate signaling pathways. We calculate the RSCU similarity score for all pairs of host-viral candidate proteins to build the network. We further analyze the host PPI network to list highly connected host proteins and highlighted a few potential drugs targeting those proteins for possible repairing of affected pathways during COVID-19.

2.1. Data acquisition and processing

Structurally, SARS-CoV-2 consists of three categories of proteins, structural, nonstructural, and accessory proteins. We select four (04) structural proteins, sixteen (16) non-structural proteins, and six (06) accessory proteins reported in [49, 50]. The details of the viral proteins
are listed in Table 1 (NCBI accession numbers for SARS-CoV-2 proteins: MN908947.3, NC_045512).

| Protein category | Count | Protein Name                  |
|------------------|-------|-------------------------------|
| Structural       | 4     | Spike (S), Envelope (E), Membrane (M), Nucleocapsid (N) |
| Non-structural   | 16    | Nsp1,Nsp2,...,Nsp16            |
| Accessory        | 6     | Orf3a, Orf6, Orf7a, Orf7b, Orf8, Orf10 |

As discussed before, we consider seventeen (17) signaling pathways, namely TGF-beta, Jak-STAT, PI3K-Akt, MAPK, HIF-1, TNF, NF-kappa B, Cytokine-cytokine receptor interaction, Apoptosis, Th17 cell differentiation, Chemokine, Toll-like receptor, RIG-like receptor, IL-17, Insulin Signaling, mTOR, and Adipocytokine, which are reported to associate with COVID-19 and other viral diseases [10, 11, 12, 13, 14, 15, 51, 52, 53]. We search the Kyoto Encyclopedia of Genes and Genomes (KEGG) database 1 for the set of human host proteins (genes) participated in our selected candidate pathways. We observe a total of 2600 genes involved in the above pathways (Supplementary-A, Table S1). We use 1313 unique genes participating exclusively in our 17 target pathways (Supplementary-B). Our proposed scheme relaying on the codon usage pattern, for which the nucleotide sequence (coding region) of each host protein is obtained from the NCBI database. A good number of genes (total 1274) are also involved in more than one pathways. We summarize our target pathways and the number of genes involved in each pathway in the Table 2.

Table 2: Candidate signaling pathways and the number of host proteins (or genes) participating in the pathway.

| Pathways                        | #Genes involved |
|---------------------------------|-----------------|
| NF-kappa B signaling pathway    | 105             |
| Cytokine-cytokine receptor interaction | 295            |
| TNF signaling pathway           | 113             |
| IL-17 signaling pathway         | 95              |
| RIG-I-like receptor signaling pathway | 70              |
| MAPK signaling pathway          | 190             |
| Chemokine signaling pathway     | 295             |
| PEK-Akt signaling pathway       | 355             |
| Jak-STAT signaling pathway      | 163             |

| Pathways                        | #Genes involved |
|---------------------------------|-----------------|
| Th17 cell differentiation       | 108             |
| TGF-beta signaling pathway      | 95              |
| Toll-like receptor signaling pathway | 105            |
| HIF-1 signaling pathway         | 110             |
| Apoptosis                       | 137             |
| Insulin signaling pathway       | 138             |
| mTOR signaling pathway          | 156             |
| Adipocytokine signaling pathway | 70              |

2.2. Computing Relative Synonymous Codon Usage (RSCU)

The genetic code describes how the 64-nucleotide triplets specify only twenty (20) different translated amino acids. These alternative codons for the same amino acids are termed as synonymous codons. However, most of the amino acids have at least two synonymous codons that are not used at the same frequencies in different genomes. Differences in the frequency of occurrence on synonymous codons in coding DNA is termed as synonymous codon usage bias [54].

1www.genome.jp/kegg/pathway.html
Several indices are available to quantify the synonymous codon usage bias [55]. Effective Number of Codons (ENC) focuses on GC content, Rare Codon (RC) focuses on the abundance of low usage codon, and Codon Adaptation Index (CAI) calculates the frequency of the overall codons based on a reference set. They are either partially capturing the usage or generating values based on the relative whole reference set. We are looking for a normalized frequency of codon usage for comparing the variation of usage between host and viral proteins. RSCU is one of the indices for measuring codon bias that is used to examine synonymous codon usage without the confounding influence of the amino acid composition of different gene products [55]. It is widely used to estimate the codon usage bias [56, 57, 58, 59]. It can be used to quantify the similarity between any two gene sequences by applying any classical proximity measure between a pair of RSCU vectors. The similarity between RSCU vectors may reflect the possible interactions between a couple of proteins in the PPI [58, 60, 57].

RSCU is the ratio between the observed number of occurrences of codons and expected during uniform usage of synonymous codons and can be calculated as follows.

\[
RSCU_{i,j} = \frac{X_{i,j}}{\frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}},
\]

where, \(X_{i,j}\) is the number of occurrences of the \(j^{th}\) codon for the \(i^{th}\) amino acid, which is encoded by \(n_i\) synonymous codons. The RSCU score of a codon more than 1.0 indicates excess usage (biased) of the codon, and less than 1.0 marks low usage of that particular codon.

We generate a 59-dimensional RSCU feature vector for each coding protein. We consider the usage pattern of only 59 codons (out of 64 available codons). We ignore 03 stop codons and uniquely coded codons ATG and TGG coded for Met and Trp amino acids, respectively [61]. For RSCU calculation, we use CAI package [62] available free at 2. Using the feature vectors, we try to draw the similarity between host and viral proteins to form a network, as discussed next.

2.3. Inferring Host-Viral Protein Interaction Network

Protein-Protein Interactions (PPI) are usually studied computationally from a graph-theoretic perspective [63]. Interactions among different organisms, such as a host and its pathogen, are primarily driven by interactions among the host proteins and pathogen proteins. These interactions can also be represented as host-pathogen PPI. Host-pathogen PPI usually represented as a bipartite graph where any given interacting pair of nodes (proteins) does not belong to the same organism. This network essentially provides the known interactions of host proteins with pathogen proteins.

Pearson’s correlation coefficient (\(\rho\)) is used to calculate relationship between two variables with different magnitudes [64, 65]. Assume \(\mathcal{R}_v = \{x_1, x_2, \cdots, x_{m=59}\}\) and \(\mathcal{R}_h = \{y_1, x_2, \cdots, y_{m=59}\}\) are the RSCU vectors for a pair of viral and host proteins, respectively. Based on \(\mathcal{R}_v\) and \(\mathcal{R}_h\),

https://cai.readthedocs.io/en/latest/
\( \rho \) can be calculated as follows.

\[
\rho(R_v, R_h) = \frac{\sum_{i=1}^{m} (x_i - \bar{R}_v)(y_i - \bar{R}_h)}{\sqrt{\sum_{i=1}^{m} (x_i - \bar{R}_v)^2} \sqrt{\sum_{i=1}^{m} (y_i - \bar{R}_h)^2}}
\]  

(2)

where, \( x_i \in R_v \) and \( y_i \in R_h \), \( \bar{R}_v \) and \( \bar{R}_h \) are the mean of the vectors \( R_v \) and \( R_h \) respectively.

To determine significantly correlated pair of RSCU vectors, we use 2-tailed \( p \) measurement \([66]\). Two proteins are strongly connected if the \( p \) is less than certain cutoff threshold, \( \tau \), i.e. \( p(R_v, R_h) < \tau \). We use SciPy version 1.5.0 \((\text{scipy.stats})^3\) for calculating \( \rho \) and \( p \) value.

We consider two \((02)\) kinds of interactions, positive and negative, between a host and viral proteins while inferring the network. Positive interaction indicates possible similar codon usage, whereas negative score signifies possible rare codon usage by SARS-CoV-2 proteins compared to its interacting host proteins. The Pearson correlation coefficient \( (\rho) \) is used to determine the possible sign of the inferred edges. An example is shown in Figure 1 for negative and positive correlations computed between two pairs of host-viral protein.

Figure 1: The host-viral codon usage (RSCU) patterns. The scatter plot shows the RSCU values of 59 codons for viral (X-axis) and host (Y-axis) proteins. The regression line represents the trends of codon RSCU patterns. Viral protein S showed positive correlation \( (\rho = 0.82) \) with host protein TANK and negative correlation \( (\rho = -0.71) \) with host protein GDF15.

Given a set of viral proteins, \( V = \{v_1, v_2, \ldots v_n\} \) and host proteins \( H = \{h_1, h_2, \ldots h_n\} \) we can create a bipartite graph in the form of adjacency matrix using above \( \rho \) and \( p \) measures as follows.

\[
I_{(V_i,H_j)} = \begin{cases} 
+1, & \text{if } p(R_{v_i}, R_{h_j}) < \tau \text{ and } \rho(R_{v_i}, R_{h_j}) > 0 \\
-1, & \text{if } p(R_{v_i}, R_{h_j}) < \tau \text{ and } \rho(R_{v_i}, R_{h_j}) < 0 \\
0, & \text{if } p(R_{v_i}, R_{h_j}) > \tau 
\end{cases}
\]  

(3)

---

[^3]: https://scipy.org
Next, we investigate the interaction mechanism of SARS-CoV-2 proteins with the proteins involved in certain signaling pathways as discussed (Table 2).

3. Results and Discussion

3.1. Benchmarking

To assess the effectiveness of codon usage bias measure in predicting possible viral protein and host protein interactions, the reported host-viral networks with physically verified interactions are considered. For example, 332 host protein interactions with 27 SARS-CoV-2 proteins [3] (network-1) are reported that utilized affinity purification mass spectrometry (AP-MS) based method to infer the physical interactions. It reports host-viral interactions forming star-like topology, where one host is exclusively interacting with one viral node. Few other similar studies report SARS-CoV-2 viral proteins interactions with more than 1100 [67](network-2) and 200 [16](network-3) host proteins. Altogether, there are total of 294 interactions (network-1), 1106 interactions (network-2), and 517 interactions (network-3) in the above networks, which are available in BioGRID database [68]. We apply our method to these three networks, and we observe approximately 54% interactions (average) for the above three networks (Table 3). In addition, we use three (03) other viral-host networks, such as Epstein-Barr from Virhostome, Hepatitis-C and Influenza-A from VirusMINT [69] for validation. We report the performance in Table 3. For more detail results one can refer to Supplementary-c.

Unlike reported physical interactions methods, codon usage infers a possible co-expression between a pair of host and viral proteins computed quantitatively using pairwise RSCU score similarity. Possibly this might be a possible reason for inferring low matching interactions with the true networks. We consider $\tau = 0.05$ for inferring the above networks.

| Virus          | Source              | # Viral proteins | # Host proteins | # Host-virus interactions | Matching (%) |
|---------------|---------------------|------------------|-----------------|---------------------------|--------------|
| SARS-CoV-2    | Network-1 [3]       | 23               | 294             | 294                       | 60           |
|               | Network-2 [67]      | 21               | 884             | 1106                      | 43           |
|               | Network-3 [16]      | 25               | 203             | 517                       | 59           |
| Epstein-Barr  | Virhostome $^4$     | 19               | 435             | 1486                      | 60           |
| Hepatitis-C   | VirusMINT [69]      | 2                | 15              | 15                        | 85           |
| Influenza-A   | VirusMINT [69]      | 2                | 4               | 4                         | 75           |

3.2. Comparison of RSCU patterns among viral and host proteins

We report codon usage distribution of 59 codons across 26 viral proteins (SARS-CoV-2) and 1313 host proteins in Figure 2 (a) and (b), respectively, involved in our candidate signaling pathways. We observe that GGT, AGA, GCT, CCT, GTT, TCT, ACA, CTT, TTA, ACT are the highly used (median RSCU score $\geq 1.5$ for each codon) codons in SARS-CoV-2 proteins. On the other hand, CGA, AGC, ACC, CGG, CTG, CCG, ACG, GCG, TCG, GGG rarely

$^4$http://interactome.dfci.harvard.edu/V_hostome/idx.php
used codons. In the host proteins (from 17 signaling pathways), codons such as CTG, GTG, ATC, GCC, CAG, ACC, AGC, GGC, and CCC are highly used (median RSCU score $\geq 1.5$ for each codon). The distribution margins of RSCU values of those codons are relatively wider (Figure 2 (b)). However, CCG, GTT, CGT, GCG, TCG, CAA, CTA, ATA, GTA, TTA rarely used codons in host proteins. It is worth mentioning that for SARS-CoV-2 proteins, highly used codons are ending (third position of codon) with T or A that shows similar characteristics with Nipah virus [54], SARS-CoV [70], and coronavirus N genes [71]. But for host proteins from candidate signaling pathways, the highly used codons are ending with G or C at the third position of the codons.

![Figure 2](image.png)

Figure 2: Distribution of RSCU scores for 59 codons for all (a) SARS-CoV-2 proteins; (b) Host proteins.

### 3.3. Analysis of host-viral inferred networks

We predict the host-viral (SARS-CoV-2) interaction graph based on the Equation 3 involving 26 SARS-CoV-2 proteins with 1313 host proteins participating in 17 signaling pathways. Out of 34138 ($26 \times 1313$) maximum possible interactions, our method infers 9412 ($\approx 36\%$) strong interactions. In our network, 859 distinct host proteins ($\approx 66\%$) are connected to at
least one viral protein. We set $\tau = 0.001$ for deciding the strong interaction (edge) between two proteins. Interestingly, our inferred network reveals that out of 859 host proteins, a total of 779 proteins is targeted by more than one viral protein. A snapshot of isolated networks with one (viral) to many (host) interactions are shown in Figure 3 between viral and host proteins.

Figure 3: The host-viral interactions network showing host proteins, which are connected to a single viral protein. In the network, the yellow-color represents viral nodes, whereas the blue and green colors represent host nodes, represent positive and negative interactions, respectively. As shown in the figure, 09 viral proteins (Nsp1, Nsp2, Nsp6, Nsp7, Nsp9, M, N, Orf3a, and Orf7a), 08 viral proteins (Nsp3, Nsp5, Nsp10, Nsp11, Nsp12, Nsp16, Orf8, and Orf10), and 03 viral proteins (Nsp13, Nsp14, and Nsp15) interactions with host proteins are positive, negative and both, respectively.

Similar researches on SARS-CoV-2 host protein interactions [3] shows viral protein oriented star-like topology only and unable to report any host protein oriented multiple interactions. We report a list of such highly connected host proteins with the viral proteins (total of 15) in Table 4. Many (viral) to one (host) interactions are also reported (Supplementary-D).
Table 4: The list of top few host proteins targeted by number of SARS-CoV-2 interacting viral proteins. For each host protein (Hp), number of interacting viral proteins (IVP) count and calculated average correlation value (ρ) are shown. There are total of 40 host proteins (20 for positive interactions and 20 for negative interactions).

| Sl. No. | Hp          | IVP count | Avg. (ρ) | SARS-CoV-2 interacting viral proteins                                                                 |
|---------|-------------|-----------|----------|-------------------------------------------------------------------------------------------------------|
| 1       | COL4A5      | 21        | 0.62     | Nsp1, Nsp2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 2       | STAM2       | 21        | 0.62     | Nsp2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 3       | LIFR        | 21        | 0.64     | Nsp2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 4       | IFNAR1      | 20        | 0.59     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 5       | PPMP        | 20        | 0.61     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 6       | RPS6KA6     | 20        | 0.62     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 7       | SOS2        | 20        | 0.63     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 8       | FK2N        | 20        | 0.66     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 9       | IRAK4       | 20        | 0.69     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 10      | IL13RA2     | 19        | 0.61     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 11      | APAF1       | 19        | 0.61     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 12      | CUL2        | 19        | 0.61     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 13      | DNML        | 19        | 0.63     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 14      | MIOS        | 19        | 0.64     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 15      | BIRC2       | 19        | 0.65     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 16      | RPS6KA3     | 19        | 0.68     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 17      | PPP113A     | 19        | 0.70     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 18      | SGK3        | 18        | 0.61     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 19      | PPP3CB      | 18        | 0.62     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
| 20      | HIF1A       | 18        | 0.62     | NpA2, NpA3, NpA4, NpA5, NpA6, NpA7, NpA8, NpA9, NpA10, NpA12, NpA13, NpA14, NpA15, NpA16, S, M, N, Orf3a, Orf7a, Orf8 |
Table 4 – continued from previous page

| Sl. No. | Hp  | Vp count | Avg. (ρ)   | Interacting viral proteins               |
|--------|-----|----------|------------|-----------------------------------------|
| 20     | FZD8 | 18       | -0.56      | Nsp2, Nsp3, Nsp4, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp10, Nsp12, Nsp13, Nsp14, Nsp15, Nsp16, S, Orf3a, Orf6, Orf8 |

3.4. Distribution of correlation scores

Statistically, it is also important to study the distribution of correlation values (both positive and negative) between pairs of proteins in terms of codon usage patterns. From the distribution plot given in Figure 4 reveals that the host-viral codon usage pattern (edge correlation) shows non-normal distribution pattern (with $p = 1.13e - 42$ for positive correlation, and $p = 7.819e - 94$ for negative correlation based on normality test performed using SciPy.stats.normaltest 3) [72, 73]. The negative correlation is varied in the range [-.73, -4.18], which covered 6325 (67%) interactions, and a positive correlation is varied in the range [4.18, 8.44], which covered 3087 (33%) interactions. So, positive correlation exhibits a wider range of values than the negative range.

We further look into the correlation value distribution of a viral protein interacting with its target proteins. We report the correlation value range (both positive and negative) for 26 viral proteins in Figure 4. While fixing $\tau$ at high (significance level) value, correlation values also appear to be significant which are ranging between ±0.05 and above. Except few, most of the viral proteins are interacting in the network both positively and negatively. Viral proteins, Orf10 and Nsp10 are interacting with their target negatively. Similarly, viral proteins like N, Nsp1 and Orf7b are interacting positively.

Based on correlation analysis, we may confirm that while a viral protein targets its a host, it mimics similar codon usage as its target to uphold the expression of target host proteins. Similarly, viral proteins use a set of codons that are rarely used in their targets to down-regulate the expression of its target. We observe that in the case of host proteins involved in signaling pathways, the majority of SARS-CoV-2 proteins aimed to break down the normal pathways by down regulating the key proteins involved in such pathways.

3.5. Degree distribution of host and viral proteins

In any interacting network, the node’s degree conveys essential information about the node’s influence within the network. In the case of host-viral PPI, a high degree viral protein (highly connected) may be a critical protein that influences the function activities a great number of host proteins. Pharmacologically, identifying such (hub) proteins may help design a small molecule that may bind with it to inhibit its influence during disease pathogenesis. The same may be applicable to host proteins. If host proteins have a high degree, it indicates that more viral proteins target the host proteins. However, it may require further investigation about its importance in its network, i.e., host-host protein networks. If a host protein is significant concerning its degree, suitable repurposed drug molecules may be identified for the same.

While focusing on highly interacting viral proteins, interestingly, we observe that the maximum number of highly interacting proteins belongs to the non-structural family. In the case
of structural proteins, S is a highly interacting (more than 600) protein. Out of accessory proteins, Orf8 shows a maximum interaction count next to protein S.

We report the degree distribution for each of the viral proteins from our network in Figure 5 (a). From the figure, it can be observed that majority of the viral proteins carrying a high node degree. Out of all the SARS-CoV-2 proteins, Nsp3 shows the maximum degree (≈ 700), which interacts with more than 80% of the candidate host proteins involved in 17 different signaling pathways. Concerning negative edges, i.e., connected, negatively, Nsp3 is still on top, followed by Nsp16, Nsp13, and few others. While considering positive edges, S, Nsp6, and Orf7a are found to be highly interactive. Few viral proteins like Nsp11, Orf7b, E, Nsp1, Orf10,
and Nsp11 are comparatively less interactive.

![Figure 5](image_url)

Figure 5: (a) The bar chart represents the host protein count for each viral protein based on correlation analysis (p-value < 0.001). Pc-positive count, Nc-negative count, positive and negative count are based on positive and negative correlations. (b) Degree distribution of 859 host proteins in terms of number of associated viral proteins (degree) count (x-axis) with host protein frequency (y-axis).

We show the degree distribution of 859 host proteins in Figure 5 (b), interacting with 26 viral proteins. From the distribution plot, we can observe that majority (82) of the host proteins are connected with only one viral node. While considering highly targeted host proteins by multiple viral proteins, we see less than ten (10) proteins are highly connected proteins with the degree 21 (maximum within the network). Even though our network is a bipartite graph, we observe that the number of low-degree nodes is high and high degree nodes are relatively low in the graph. It further indicates that hub or central nodes are relatively less, which is somehow following the scale-free properties [74] of a complex network. We observe relatively good host nodes possessing a degree within the range of 11 to 18.

3.6. Ranking highly targeted signaling pathways and its impact on COVID-19

To study the most affected pathways in our 17 candidate set of pathways, we rank them based on the percentage of host proteins targeted by any viral proteins out of total proteins involved in those pathways, and report in Figure 6.

The topmost pathway is Mitogen-Activated Protein Kinase (MAPK) signaling pathway. More than 50% proteins from this pathway are targeted by different SARS-CoV-2 proteins. This pathway is associated with the COVID-19 immune response [75], and involving in papain-like protease activation of promoter as observed in SARS coronavirus [76]. MAPK proteins communicate signals from a receptor on the cell’s surface to the DNA in the cell’s nucleus, which is essential in a viral infection point of view. Further, MAPK proteins are involved in a series of vital signal transduction pathways that regulate processes such as cell proliferation, cell differentiation, and cell death in humans.

Besides MAPK, other ranked signaling pathways are significantly affected during COVID-19 infection. Under physiological conditions, adipokines act mainly in adipose tissue (paracrine
or autocrine) or circulate through the blood circulation to distant target organs, regulating their growth and development, metabolism, and tissue remodeling. However, under pathological conditions, adipokines’ synthesis and secretion are disordered, leading to obesity, diabetes, heart disease, and other metabolic disorders. Our results show that the adipocytokine pathway is affected by COVID-19. It implicates that patients with comorbid conditions like diabetes and heart disease may show worst disease aggression, which is already observed in various reports.

The mTOR pathway is a central regulator of mammalian metabolism and physiology, with essential roles in tissues’ function, including liver, muscle, white and brown adipose tissue, and the brain. It is dysregulated in human diseases, such as diabetes, obesity, depression, aging-related problems, and certain cancers. Our result corroborates with the same, and it’s reported that aged patients are more prone to the infection due to the dysregulation of the m-TOR pathway or some other unknown reasons.

It has been observed that some COVID-19 affected deaths are due to multiple organ failure. HIF1 and RIG1 like receptor pathways are involved in normal immunoregulation and various organ functioning. Dysregulation may cause immune compromisation and multiple organ failure through ischaemic heart disease, acute lung injury, pulmonary hypertension, pulmonary fibrosis, chronic obstructive pulmonary disease (COPD), acute liver failure, liver fibrosis, and acute kidney injury, etc. Our result also supports these findings.

In our ranking, the fourth most affected pathway is the TGF-β (Transforming growth factor-beta), which is a multi-functional cytokine belonging to the transforming growth factor superfamily that includes three different mammalian isoforms (TGF-β 1 to 3, HGNC symbols TGFBI, TGFB2, TGFB3) and many other signaling proteins. All-white blood cell lineages produce TGFB proteins. This pathway activates different downstream substrates and regulatory proteins, inducing transcription of various target genes that function in differentiation, chemotaxis, proliferation, and activation of many immune cells.

Figure 6: Ranking of 17 candidate signaling pathways. The pathway ranking is done by observing the host protein percentage from pathways that interact with any of the SARS-CoV-2 (26) proteins.
3.7. Centrality analysis of targeted host proteins and candidate signaling pathways

Studies on human host-viral protein interactions reveal that virus tending to targeted attacks towards host proteins [77, 19, 78] by interacting with key (central) host proteins. We consider a host protein important if it interacts with many other host proteins in host-host protein networks. We use BioGRID [68] to calculate the centrality score of our candidate host proteins \(^5\). We report the top 100 central proteins in the Supplementary-A (Table S2). We observe that a good number of interacting host proteins in our network are highly central in their own (host) PPI. A common set of viral proteins targets central genes, and such proteins are involved in multiple pathways. For instance, if we consider few top central proteins, MYC (2843), TRIM25 (2656), EGFR (2452), BRCA1 (2236), MDM2 (2219), NTRK1 (2030), KRAS (1944), ELAVL1 (1914) and HSP90AA1 (1734), they are found to be targeted jointly by the viral proteins such Nsp2, Nsp3, Nsp4, Nsp5, Nsp8, Nsp10, Nsp12, Nsp13.

If we consider most central proteins in our candidate pathways, we observe that the PI3K-Akt signaling pathway (36 targeted nodes) and MAPK signaling pathway (35 targeted nodes) contains most of the central proteins targeted by the viral proteins. A pathway may be more crucial from the disease pathogenesis perspective if it contains highly central proteins targeted by viral proteins. Moving one step ahead, we may rank our 17 pathways based on the number of participating central proteins (out of top 100 centrality list) in the above pathways and shown in Figure 7(a). More details about the top 100 central host proteins are listed in Supplementary-A, (Table S2). Interestingly, in terms of the number of targeted proteins, which are also central in host-host PPI, the signaling pathway MAPK is one of the worst affected pathways among 17 candidate pathways. In addition to PPI centrality, we study the pathway centrality of the host proteins regarding our 17 signaling pathways. Prior researches also identified an exciting fact that viral proteins target host proteins that are pathway central, i.e., participating in multiple pathways [78]. Degree distribution of host proteins in terms of their density of participation in 17 pathways is reported in Figure 7(b). We observe a nice power-law [74] like distribution where the majority of proteins are participating in only one pathway, and fewer numbers are having high participation in multiple pathways. We list a few top highly pathway-central proteins and few interesting facts in Table 5. The table shows that the pathway-central proteins are also highly connected in their own PPI and mostly targeted by multiple viral proteins.

3.8. Quantitative association of key pathway proteins and drugs

To investigate further the significance of key proteins in our network, we analyse protein-drug association. We primarily consider approved or under trial drugs that are in use during COVID-19. We searched online drug target resource database \(^6\) to count hits with different key proteins in our network (Supplementary-A, Table S3). We ranked those drugs based on their counts of protein targets in our network (Table 6). A good number of drugs are also observed to be associated with central proteins that are not reported so far used in COVID-19. The list

---

\(^5\)https://thebiogrid.org/
\(^6\)http://www.dgidb.org/
Figure 7: (a) Participation host protein count of central proteins in candidate pathways; (b) Degree distribution of 859 interacting host proteins in terms of number of associated signaling pathways (candidate).

Table 5: Few top pathway central proteins with the number of pathways they participating (out of 17 pathways), PPI centrality score and number of viral proteins (Vp) targeting the proteins

| Host protein | #Pathway centrality | PPI centrality | Interacting Vp count |
|--------------|---------------------|----------------|----------------------|
| IKBKB        | 13                  | 552            | 2                    |
| CHUK         | 12                  | 462            | 11                   |
| MAPK3        | 12                  | 337            | 16                   |
| RELA         | 12                  | 859            | 10                   |
| AKT1         | 11                  | 886            | 11                   |
| AKT2         | 11                  | 113            | 12                   |
| AKT3         | 11                  | 61             | 15                   |
| IKBKG        | 11                  | 959            | 9                    |
| TNP          | 11                  | 497            | 11                   |
| MAPK8        | 9                   | 444            | 12                   |
| MAPK9        | 9                   | 260            | 15                   |
| NFKBIA       | 9                   | 504            | 8                    |
| PIK3CA       | 9                   | 190            | 19                   |
| PIK3CB       | 9                   | 82             | 17                   |
| PIK3CD       | 9                   | 28             | 17                   |
| PIK3R1       | 9                   | 684            | 5                    |
| PIK3R2       | 9                   | 190            | 16                   |

of such drugs are given in Supplementary-A (Table S3). It can be observed from the Table 6 that a single drug is having targets in multiple pathways forming a bipartite graph as shown in Figure 8.

We discuss below the drugs that are associated with COVID-19 disease and having potential target host proteins involved in our inferred host-viral networks.

**Arsenic trioxide** is a widely known chemical used for multiple disease condition. The Ministry of Ayush ⁷, Govt. Of India, advised for Arsenicum album 30 as a potential homeopath drugs for COVID-19. Arsenicum album 30 is a mother tincture of arsenic trioxide used as a homeopath medicine. Symptoms like severe respiratory adverse effects frequently

⁷[www.ayush.gov.in](https://www.ayush.gov.in)
occur in patients with promyelocytic leukemia. Arsenic trioxide could be used in consolidation therapy [79, 80]. If we consider central genes (present in the top 100 list) involved in MAPK pathways, we observe six target proteins (RARA, FGFR1, IKBKB, CCND1, CDKN1A, JUN, MAPK3, AKT1) are the good target of this chemical. Interestingly, all such proteins are targeted negatively by viral proteins. In a comorbid situation where these signaling pathway genes are already perturbed, arsenic trioxide may play a protective role in boosting up the immunity and other unknown vital regulators that are yet to discover. In addition to MAPK, several targets are present in PI3K-Akt, TNF, and Apoptosis signaling pathways.

**Dexamethasone** is another most widely used COVID-19 drug with 64 target genes. This is the first drug to show life-saving efficacy in patients infected with COVID-19 [81], and widely utilized in a large trial in the UK [82] Our result shows that NTRK1, HSPA8, SMAD3, VCAM1, and RARA are the targets (central) for the drug involved in MAPK, PI3K-Akt, Th17 cell differentiation, TGF-beta, and NF-kappa B. In addition to that, Dexamethasone also targets a few other interacting host proteins (low centrality), JUNB, LIF, CD86, SLC2A4, and IRS2. Dexamethasone is predicted to maintain these signaling pathways normal functioning and shows protection against COVID-19 symptoms, as we assume from our results.

**Hydroxychloroquine** is another important drug, have been widely utilized for COVID-19 treatment [83, 84]. The only central target is TNF, which is present in several signaling pathways. It can rapidly be transcribed in a variety of cell types following exposure to a broad range of pathogens and signals of inflammation and stress [85]. Other low centrality target are TLR3, TLR7, PTGS2, and TLR9.

Other two important drugs recommended by WHO are **Ritonavir** and **Interferon Alfa B** observed in our list used for COVID-19 trail [86, 87, 88]. Interestingly, we found that ritonavir shows three target central genes (CXCL10, TLR4, IFNL3) in our study. These genes share NF kappa B, HIF1, toll-like receptor, PI3K-AKT, JAK-STAT, cytokine-cytokine receptor, TNF, IL7, RIG1 receptor, chemokine signaling pathways. **Interferon Alfa 1B** is another option for solidarity trial which having targets like IFNAR1, IFNAR2, and IL13 genes. These genes participates in PI3K-AKT, toll-like receptor, cytokine-cytokine receptor, Jak-STAT, Il17 pathways. These pathways are essential for maintaining normal immunological functioning, which is thought to dysregulate in COVID-19.

We believe that a pathway that is mostly targeted by different SARS-CoV-2 proteins and involves highly central proteins in its PPI is the most crucial (affected) pathway. It is worth to mention that the above highlighted drugs molecules are based on quantitative analysis of host proteins from our inferred host-viral network and their hits with the existing drug target database. Hence, our proposed scheme is not a new drug repurposing methodology and needs to put more serious attention while designing therapeutic solutions pharmacologically.

---

8http://www.dgidb.org/
9www.who.int
Table 6: Few COVID-19 drugs with their actual number of target host proteins from our inferred network, number of targets that are highly central and number of targets involved in candidate pathways.

| Drug Name          | #Actual targets | #Targets in inferred networks | #Targets (central) in top-100 host PPI | #Involved pathway |
|--------------------|-----------------|------------------------------|----------------------------------------|-------------------|
| Arsenic trioxide   | 25              | 11                           | 8                                      | 16                |
| Dexamethasone      | 64              | 10                           | 5                                      | 12                |
| Hydroxychloroquine | 9               | 5                            | 1                                      | 11                |
| Interferon beta    | 5               | 4                            | 0                                      | 9                 |
| Ritonavir          | 15              | 3                            | 0                                      | 10                |

Figure 8: Bipartite graph showing 19 drugs linked with 17 signaling pathways. Left and right panel are showing drug name and middle panel is showing signaling pathway name.

4. Conclusion

This work puts a novel effort into recreating host-viral PPI. Proposed work explored the codon usage pattern similarity between coding DNA sequences of host proteins participating in a few major signaling pathways and SARS-CoV-2 viral proteins. Both positive and negative edges between interacting proteins were inferred, which depict an essential association between viral and host proteins. The inferred network was analyzed topologically, considering nodes’ degree distribution and node centrality. An interesting fact has been observed on how viral proteins are targeting their host proteins. Our analysis highlighted a few drugs, already in use for COVID-19, having potential targets in some of the essential host proteins involved in important candidate signaling pathways such as MAPK and PI3-Akt. Several central proteins were identified (AKT1, CCND1, CDKN1A, FGFR1, HSPA8, IKBKB, JUN, MAPK3, NTRK1, RARA, SMAD3, TNF, and VCAM1), which are involved in critical signaling pathways and targeted by few drug molecules. The topmost few drug molecules highlighted by this study are Arsenic trioxide, Dexamethasone and Hydroxychloroquine, which might play an influential
role in preventing COVID-19 mortality.

Our method is generic and useful to draw a more extensive network, covering genes from all critical pathways. Even the method can be applied to any set of host-viral proteins (other than SARS-CoV-2 or Human). Currently, we used a correlation score to measure the similarity between two RSCU vectors. However, other measures such as cosine similarity, mutual information, and ensemble approach might improve the result. One may consider a multi-layer network approach considering viral-viral and host-host networks that may shed better light on the possible viral-host interaction patterns.

References

[1] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, Cryo-em structure of the 2019-ncov spike in the prefusion conformation, Science 367 (2020) 1260–1263.

[2] U. Jain, Effect of covid-19 on the organs, Cureus 12 (2020).

[3] D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O’Meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney, et al., A sars-cov-2 protein interaction map reveals targets for drug repurposing, Nature (2020) 1–13.

[4] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F. Cheng, Network-based drug repurposing for novel coronavirus 2019-ncov/sars-cov-2, Cell discovery 6 (2020) 1–18.

[5] B. R. Beck, B. Shin, Y. Choi, S. Park, K. Kang, Predicting commercially available antiviral drugs that may act on the novel coronavirus (sars-cov-2) through a drug-target interaction deep learning model, Computational and structural biotechnology journal (2020).

[6] V. Memišević, N. Zavaljevski, S. V. Rajagopala, K. Kwon, R. Pieper, D. DeShazer, J. Reifman, A. Wallqvist, Mining host-pathogen protein interactions to characterize burkholderia mallei infectivity mechanisms, PLoS Comput Biol 11 (2015) e1004088.

[7] A. Pichlmair, K. Kandasamy, G. Alvisi, O. Mulhern, R. Sacco, M. Habjan, M. Binder, A. Stefanovic, C.-A. Eberle, A. Goncalves, et al., Viral immune modulators perturb the human molecular network by common and unique strategies, Nature 487 (2012) 486–490.

[8] R. Draenert, J. Frater, J. G. Prado, Virus immune evasion: new mechanism and implications in disease outcome, 2012.

[9] A. Kieser, Signal transduction by viral factors: critical interface between the virus and its host cell with implications for the viral life cycle and disease development, Signal Transduction 7 (2007) 3–4.

[10] F. Seif, H. Aazami, M. Khoshmirsafa, M. Kamali, M. Mohsenzadegan, M. Pornour, D. Mansouri, Jak inhibition as a new treatment strategy for patients with covid-19, International Archives of Allergy and Immunology 181 (2020) 467–475.
[11] D. Wu, X. O. Yang, Th17 responses in cytokine storm of covid-19: An emerging target of jak2 inhibitor fedratinib, Journal of Microbiology, Immunology and Infection 53 (2020) 368–370.

[12] H. Ganesan, V. Balasubramanian, M. Iyer, A. Venugopal, M. D. Subramaniam, S.-G. Cho, B. Vellingiri, mtor signalling pathway-a root cause for idiopathic autism?, BMB reports 52 (2019) 424.

[13] W. Luo, Y.-X. Li, L.-J. Jiang, Q. Chen, T. Wang, D.-W. Ye, Targeting jak-stat signaling to control cytokine release syndrome in covid-19, Trends in pharmacological sciences (2020).

[14] T. Shibabaw, Inflammatory cytokine: Il-17a signaling pathway in patients present with covid-19 and current treatment strategy, Journal of inflammation research 13 (2020) 673.

[15] J. M. Grimes, K. V. Grimes, p38 mapk inhibition: A promising therapeutic approach for covid-19, Journal of Molecular and Cellular Cardiology (2020).

[16] J. Li, M. Guo, X. Tian, X. Wang, X. Yang, P. Wu, C. Liu, Z. Xiao, Y. Qu, Y. Yin, et al., Virus-host interactome and proteomic survey of pbmcs from covid-19 patients reveal potential virulence factors influencing sars-cov-2 pathogenesis, Med (2020). doi:10.1016/j.medj.2020.07.002.

[17] F. Messina, E. Giombini, C. Agrati, F. Vairo, T. A. Bartoli, S. Al Moghazi, M. Piacentini, F. Locatelli, G. Kobinger, M. Maeurer, et al., Covid-19: viral–host interactome analyzed by network based-approach model to study pathogenesis of sars-cov-2 infection, Journal of Translational Medicine 18 (2020) 1–10.

[18] J. Kumar Das, G. Tradigo, P. Veltri, P. H Guzzi, S. Roy, Data science in unveiling covid-19 pathogenesis and diagnosis: evolutionary origin to drug repurposing, Briefings in Bioinformatics 22 (2021) 855–872.

[19] V. Navratil, B. de Chassey, C. R. Combe, V. Lotteau, When the human viral infectome and diseasome networks collide: towards a systems biology platform for the aetiology of human diseases, BMC systems biology 5 (2011) 13.

[20] M. D. Dyer, T. Murali, B. W. Sobral, The landscape of human proteins interacting with viruses and other pathogens, PLoS Pathog 4 (2008) e32.

[21] O. Kuchaiev, M. Rašajski, D. J. Higham, N. Pržulj, Geometric de-noising of protein-protein interaction networks, PLoS computational biology 5 (2009) e1000454.

[22] Y. Murakami, K. Mizuguchi, Homology-based prediction of interactions between proteins using averaged one-dependence estimators, BMC bioinformatics 15 (2014) 213.

[23] L. Salwinski, D. Eisenberg, Computational methods of analysis of protein–protein interactions, Current opinion in structural biology 13 (2003) 377–382.
[24] Q. C. Zhang, D. Petrey, L. Deng, L. Qiang, Y. Shi, C. A. Thu, B. Bisikirksa, C. Lefebvre, D. Accili, T. Hunter, et al., Structure-based prediction of protein–protein interactions on a genome-wide scale, Nature 490 (2012) 556–560.

[25] R. Singh, D. Park, J. Xu, R. Hosur, B. Berger, Struct2net: a web service to predict protein–protein interactions using a structure-based approach, Nucleic acids research 38 (2010) W508–W515.

[26] S. Jain, G. D. Bader, An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology, BMC bioinformatics 11 (2010) 1–14.

[27] X. Yang, S. Yang, Q. Li, S. Wuchty, Z. Zhang, Prediction of human-virus protein-protein interactions through a sequence embedding-based machine learning method, Computational and structural biotechnology journal 18 (2020) 153–161.

[28] S. Alguwaizani, B. Park, X. Zhou, D.-S. Huang, K. Han, Predicting interactions between virus and host proteins using repeat patterns and composition of amino acids, Journal of healthcare engineering 2018 (2018).

[29] M. Dilucca, S. Forcelloni, A. G. Georgakilas, A. Giansanti, A. Pavlopoulou, Codon usage and phenotypic divergences of sars-cov-2 genes, Viruses 12 (2020) 498.

[30] P. M. Sharp, L. R. Emery, K. Zeng, Forces that influence the evolution of codon bias, Philosophical Transactions of the Royal Society B: Biological Sciences 365 (2010) 1203–1212.

[31] G. M. Jenkins, E. C. Holmes, The extent of codon usage bias in human rna viruses and its evolutionary origin, Virus research 92 (2003) 1–7.

[32] M. Gale, S.-L. Tan, M. G. Katze, Translational control of viral gene expression in eukaryotes, Microbiology and molecular biology reviews 64 (2000) 239–280.

[33] G. Lithwick, H. Margalit, Relative predicted protein levels of functionally associated proteins are conserved across organisms, Nucleic acids research 33 (2005) 1051–1057.

[34] H. B. Fraser, A. E. Hirsh, D. P. Wall, M. B. Eisen, Coevolution of gene expression among interacting proteins, Proceedings of the National Academy of Sciences 101 (2004) 9033–9038.

[35] I. Bahir, M. Fromer, Y. Prat, M. Linial, Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences, Molecular systems biology 5 (2009) 311.

[36] Z. Zhou, Y. Dang, M. Zhou, L. Li, C.-h. Yu, J. Fu, S. Chen, Y. Liu, Codon usage is an important determinant of gene expression levels largely through its effects on transcription, Proceedings of the National Academy of Sciences 113 (2016) E6117–E6125.
[37] K. Jitobaom, S. Phakaratsakul, T. Sirihongthong, S. Chotewutmontri, P. Suriyaphol, O. Suptawiwat, P. Auewarakul, Codon usage similarity between viral and some host genes suggests a codon-specific translational regulation, Heliyon 6 (2020) e03915.

[38] H. Song, H. Gao, J. Liu, P. Tian, Z. Nan, Comprehensive analysis of correlations among codon usage bias, gene expression, and substitution rate in arachis duranensis and arachis ipaënsis orthologs, Scientific reports 7 (2017) 1–12.

[39] M. Dos Reis, L. Wernisch, R. Savva, Unexpected correlations between gene expression and codon usage bias from microarray data for the whole escherichia coli k-12 genome, Nucleic acids research 31 (2003) 6976–6985.

[40] J. L. Chaney, P. L. Clark, Roles for synonymous codon usage in protein biogenesis, Annual review of biophysics 44 (2015) 143–166.

[41] Y. Zhou, Y.-S. Zhou, F. He, J. Song, Z. Zhang, Can simple codon pair usage predict protein–protein interaction?, Molecular BioSystems 8 (2012) 1396–1404.

[42] X. Peng, J. Wang, W. Peng, F.-X. Wu, Y. Pan, Protein–protein interactions: detection, reliability assessment and applications, Briefings in Bioinformatics 18 (2016) 798–819.

[43] V. S. Rao, K. Srinivas, G. Sujini, G. Kumar, Protein-protein interaction detection: methods and analysis, International journal of proteomics 2014 (2014).

[44] R. Grantham, C. Gautier, M. Gouy, R. Mercier, A. Pave, Codon catalog usage and the genome hypothesis, Nucleic acids research 8 (1980) 197–197.

[45] Y.-s. Liu, J.-h. Zhou, H.-t. Chen, L.-n. Ma, Z. Pejsak, Y.-z. Ding, J. Zhang, The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern, Infection, Genetics and Evolution 11 (2011) 1168–1173.

[46] J. B. Plotkin, H. Robins, A. J. Levine, Tissue-specific codon usage and the expression of human genes, Proceedings of the National Academy of Sciences 101 (2004) 12588–12591.

[47] S. Mueller, D. Papamichail, J. R. Coleman, S. Skiena, E. Wimmer, Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity, Journal of virology 80 (2006) 9687–9696.

[48] R. W. Tindle, Immune evasion in human papillomavirus-associated cervical cancer, Nature Reviews Cancer 2 (2002) 59–64.

[49] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, et al., A new coronavirus associated with human respiratory disease in china, Nature 579 (2020) 265–269.

[50] J. K. Das, S. Roy, A study on non-synonymous mutational patterns in structural proteins of sars-cov-2, Genome (2021). doi:10.1139/gen-2020-0157.
[51] M. A.-A.-K. Khan, A. B. M. M. K. Islam, Sars-cov-2 proteins exploit host’s genetic and epigenetic mediators for the annexation of key host signaling pathways, Frontiers in Molecular Biosciences 7 (2020).

[52] L. Chen, H. Deng, H. Cui, J. Fang, Z. Zuo, J. Deng, Y. Li, X. Wang, L. Zhao, Inflammatory responses and inflammation-associated diseases in organs, Oncotarget 9 (2018) 7204.

[53] Z. Zhao, J. Xia, O. Tastan, I. Singh, M. Kshirsagar, J. Carbonell, J. Klein-Seetharaman, Virus interactions with human signal transduction pathways, International journal of computational biology and drug design 4 (2011) 83–105.

[54] R. Khandia, S. Singhal, U. Kumar, A. Ansari, R. Tiwari, K. Dhama, J. Das, A. Munjal, R. K. Singh, Analysis of nipah virus codon usage and adaptation to hosts, Frontiers in microbiology 10 (2019) 886.

[55] P. M. Sharp, T. M. Tuohy, K. R. Mosurski, Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes, Nucleic acids research 14 (1986) 5125–5143.

[56] P. M. Sharp, W.-H. Li, An evolutionary perspective on synonymous codon usage in unicellular organisms, Journal of molecular evolution 24 (1986) 28–38.

[57] M. Dilucca, S. Forcelloni, G. Cimini, A. Giansanti, Co-evolution between codon usage and protein-protein interaction networks in bacterial genomes, bioRxiv (2020).

[58] H. S. Najafabadi, R. Salavati, Sequence-based prediction of protein-protein interactions by means of codon usage, Genome biology 9 (2008) 1–9.

[59] J. Das, S. Roy, Comparative analysis of human coronaviruses focusing on nucleotide variability and synonymous codon usage pattern, BioRxiv (2020).

[60] P. L. Meintjes, A. G. Rodrigo, Evolution of relative synonymous codon usage in human immunodeficiency virus type-1, Journal of bioinformatics and computational biology 3 (2005) 157–168.

[61] Y. Chen, Q. Xu, X. Yuan, X. Li, T. Zhu, Y. Ma, J.-L. Chen, Analysis of the codon usage pattern in middle east respiratory syndrome coronavirus, Oncotarget 8 (2017) 110337.

[62] B. D. Lee, Python implementation of codon adaptation index, Journal of Open Source Software 3 (2018) 905. URL: https://doi.org/10.21105/joss.00905. doi:10.21105/joss.00905.

[63] P. H. Guzzi, S. Roy, Biological Network Analysis: Trends, Approaches, Graph Theory, and Algorithms, Academic Press, 2020.
[64] B. G. Häne, K. Jäger, H. G. Drexler, The pearson product-moment correlation coefficient is better suited for identification of dna fingerprint profiles than band matching algorithms, Electrophoresis 14 (1993) 967–972.

[65] J. Benesty, J. Chen, Y. Huang, I. Cohen, Pearson correlation coefficient, in: Noise reduction in speech processing, Springer, 2009, pp. 1–4.

[66] P. Sedgwick, Pearson’s correlation coefficient, Bmj 345 (2012) e4483.

[67] A. Stukalov, V. Girault, V. Grass, V. Bergant, O. Karayel, C. Urban, D. A. Haas, Y. Huang, L. Oubraham, A. Wang, et al., Multi-level proteomics reveals host-perturbation strategies of sars-cov-2 and sars-cov, Biorxiv (2020).

[68] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, Biogrid: a general repository for interaction datasets, Nucleic acids research 34 (2006) D535–D539.

[69] A. Chatr-Aryamontri, A. Ceol, D. Peluso, A. Nardozza, S. Panni, F. Sacco, M. Tinti, A. Smolyar, L. Castagnoli, M. Vidal, et al., Virusmint: a viral protein interaction database, Nucleic acids research 37 (2009) D669–D673.

[70] W. Gu, T. Zhou, J. Ma, X. Sun, Z. Lu, Analysis of synonymous codon usage in sars coronavirus and other viruses in the nidovirales, Virus research 101 (2004) 155–161.

[71] A. Sheikh, A. Al-Taher, M. Al-Nazawi, A. I. Al-Mubarak, M. Kandeel, Analysis of preferred codon usage in the coronavirus n genes and their implications for genome evolution and vaccine design, Journal of virological methods 277 (2020) 113806.

[72] R. B. d’Agostino, An omnibus test of normality for moderate and large size samples, Biometrika 58 (1971) 341–348.

[73] R. D’AGOSTINO, E. S. Pearson, Tests for departure from normality. empirical results for the distributions of b2 and b, Biometrika 60 (1973) 613–622.

[74] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Reviews of modern physics 74 (2002) 47.

[75] L. Huang, Y. Shi, B. Gong, L. Jiang, X. Liu, J. Yang, J. Tang, C. You, Q. Jiang, B. Long, et al., Blood single cell immune profiling reveals the interferon-mapk pathway mediated adaptive immune response for covid-19, MedRxiv (2020).

[76] S.-W. Li, C.-Y. Wang, Y.-J. Jou, T.-C. Yang, S.-H. Huang, L. Wan, Y.-J. Lin, C.-W. Lin, Sars coronavirus papain-like protease induces egr-1-dependent up-regulation of tgf-β1 via ros/p38 mapk/stat3 pathway, Scientific reports 6 (2016) 1–13.

[77] R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks, nature 406 (2000) 378–382.
[78] R. R. Halehalli, H. A. Nagarajaram, Molecular principles of human virus protein–protein interactions, Bioinformatics 31 (2015) 1025–1033.

[79] M. Gavillet, J. C. Klappert, O. Spertini, S. Blum, Acute leukemia in the time of covid-19, Leukemia research 92 (2020) 106353.

[80] F. Ferrara, P. Zappasodi, E. Roncoroni, E. Borlenghi, G. Rossi, Impact of covid-19 on the treatment of acute myeloid leukemia, Leukemia 34 (2020) 2254–2256.

[81] T. Lammers, A. M. Sofias, R. van der Meel, R. Schifflers, G. Storm, F. Tacke, S. Koschmieder, T. H. Brümmendorf, F. Kiessling, J. M. Metselaar, Dexamethasone nanomedicines for covid-19, Nature nanotechnology 15 (2020) 622–624.

[82] T. Theoharides, P. Conti, Dexamethasone for covid-19? not so fast, J Biol Regul Homeost Agents 34 (2020) 10–23812.

[83] Z. Chen, J. Hu, Z. Zhang, S. Jiang, S. Han, D. Yan, R. Zhuang, B. Hu, Z. Zhang, Efficacy of hydroxychloroquine in patients with covid-19: results of a randomized clinical trial, medrxiv (2020).

[84] C. P. Skipper, K. A. Pastick, N. W. Engen, A. S. Bangdiwala, M. Abassi, S. M. Lofgren, D. A. Williams, E. C. Okafor, M. F. Pullen, M. R. Nicol, et al., Hydroxychloroquine in nonhospitalized adults with early covid-19: a randomized trial, Annals of internal medicine 173 (2020) 623–631.

[85] J. V. Falvo, A. V. Tsytsykova, A. E. Goldfeld, Transcriptional control of the tnf gene, TNF Pathophysiology 11 (2010) 27–60.

[86] P. Dalerba, B. Levin, J. L. Thompson, A trial of lopinavir-ritonavir in covid-19., The New England journal of medicine 382 (2020).

[87] X. Ye, Y. Luo, S. Xia, Q. Sun, J. Ding, Y. Zhou, W. Chen, X. Wang, W. Zhang, W. Du, et al., Clinical efficacy of lopinavir/ritonavir in the treatment of coronavirus disease 2019, Eur Rev Med Pharmacol Sci 24 (2020) 3390–3396.

[88] L. Zha, S. Li, L. Pan, B. Tefsen, Y. Li, N. French, L. Chen, G. Yang, E. V. Villanueva, Corticosteroid treatment of patients with coronavirus disease 2019 (covid-19), Medical Journal of Australia 212 (2020) 416–420.