

# Master Regulator Analysis of the SARS-CoV-2/Human interactome

Pietro Hiram Guzzi <sup>1§</sup>, Daniele Mercatelli <sup>2§</sup>, Carmine Ceraolo<sup>2</sup> and Federico M. Giorgi <sup>2,\*</sup>

<sup>1</sup> Department of Surgical and Medical Science, University of Catanzaro, Italy

<sup>2</sup> Department of Pharmacy and Biotechnology, University of Bologna, Italy

\* Correspondence: federico.giorgi@unibo.it (F.M.G.)

§ These authors contributed equally to this work

**Abstract:** the recent epidemic outbreak of a novel human coronavirus called SARS-CoV-2 and causing the respiratory tract disease COVID-19 has reached worldwide resonance and a global effort is being undertaken to characterize the molecular features and evolutionary origins of this virus. In this paper, we set out to shed light on the SARS-CoV-2/host receptor recognition, a crucial factor for successful virus infection. Based on the current knowledge of the interactome between SARS-CoV-2 and host cell proteins, we performed Master Regulator Analysis to detect which parts of the human interactome are most affected by the infection. We detected, amongst others, affected apoptotic and mitochondrial mechanisms, and a downregulation of the ACE2 protein receptor, notions that can be used to develop specific therapies against this new virus.

**Keywords:** coronavirus; bioinformatics; gene network analysis

---

## 1. Introduction

Beta-coronaviruses are a subgroup of the coronavirus family, large enveloped positive-stranded RNA viruses able to infect a wide variety of mammals and avian species, causing mainly respiratory or enteric diseases [1]. They present surface spikes formed by (S) glycoproteins which mediate binding to susceptible host cell receptors to initiate the infection [2]. To date, seven beta-coronaviruses are known to cause human disease. The prevalent strains (HCoV 229E, HKU1, NL63 and OC43) typically cause mild infections of the upper respiratory tract in humans [3], while three strains were found to pose a severe health threat, leading in some cases to a fatal outcome: SARS-CoV, MERS-CoV and the newly identified SARS-CoV-2 [4]. SARS-CoV was recognized as the causal agent of the severe acute respiratory syndrome outbreak occurring in China in 2002 and 2003 [5], while MERS-CoV was responsible for the Middle East respiratory syndrome occurring in 2012 [6]. In both cases, the infected patients manifested severe acute pneumonia, but while SARS-CoV infected mainly the lower respiratory tract, MERS-CoV caused more pronounced gastrointestinal symptoms, often associated with kidney failure [7]. This can be possibly explained by a difference in recognition of host surface receptors. While the primary human receptor for SARS S glycoprotein is the angiotensin-converting enzyme 2 (ACE2), which is widely expressed on the surface of the cells in the lower respiratory tract, MERS-CoV preferentially binds to surface enzyme dipeptidyl peptidase 4 (DPP4, also known as CD26), which is present both in the lower respiratory tract as well as kidneys and the gastrointestinal tract [8–10]. This distribution may also explain the observed difference in transmission of the two viruses [7]. The third highly pathogenic strain of coronavirus, SARS-CoV-2, appeared in late December 2019 in Wuhan, Hubei Province, China [11]. It is a highly aggressive new coronavirus which has infected more than 83,000 people (at February 29), with an overall case-fatality rate of 2.3% [12]. In less than two months, the SARS-CoV-2

outbreak spread from the Hubei Province to 26 other countries, reaching also the US and Europe, being Italy the European country with the highest number of patients tested positive for the new coronavirus [13]. Given the emergency, the characterization of the biology of SARS-CoV-2 and features of virus-host interactions is demanding to identify candidate molecules to inhibit viral functions. Sequence similarity of new coronavirus with SARS and MERS was previously reported and close similarities with wild bat coronaviruses has been reported [14]. New coronavirus spike S glycoprotein sequence shares a sequence identity above 72% with human SARS, showing a unique furin-like cleavage site which is a feature absent in the other SARS-like CoVs [15]. Biophysical and structural evidence that new SARS-CoV-2 may bind ACE2 receptor with higher affinity than SARS-CoV has been recently reported [9]. The wide distribution and similarity of this receptor in the animal kingdom may account for cross-species transmission [16], while the pattern of expression of ACE2 in human respiratory epithelia and oral mucosa may explain the fast human-human transmission [17].

Despite the crucial need for a clinical strategy to counteract this epidemiological outbreak, no specific drug has been identified yet. This will require the elucidation of genetic features, molecular constituents and mechanisms of infection, which ultimately reside in the mechanisms regulating virus-host interaction and host responses. However, there is a lack of molecular data describing the COVID-19, and those available are related to the genomic and protein sequence of this virus [14] and there is currently no high-throughput dataset describing the mechanism of infection, e.g. transcriptomics of infected lung cells.

In this context, the use of bioinformatics approaches may help filling the gap of missing data and improve the process of knowledge discovery by elucidating molecular mechanisms behind virus replication, the process of viral attachment to the host cells and the effect on the host molecular pathways [18]. The final understanding of the molecular action of the virus may improve or accelerate the development of anti-viral therapeutic approaches using information and data available for other coronaviruses [19].

Here, we applied a network approach to investigate cellular responses to the interaction of human cells with SARS and MERS viruses to get more insights into the molecular mechanisms regulating susceptibility to infection and transmission. The analysis of cell-context specific repertoires of molecular interactions, or interactomes, are becoming increasingly useful to dissect molecular pathways and biological processes, because they offer the possibility to get a comprehensive overview of cellular regulatory programs [20]. Integration of interactome maps to gene-regulatory networks (i.e. topological maps representing relationships between transcription factors and their target genes), can better provide insights into critical cellular elements that regulate a certain phenotypic state or response [21]. It has been observed that a number of large-scale transcriptional cascades behind complex cellular processes can be explained by the action of a relatively small number of transcription factors, or Master Regulators [22–24]. The identification of most crucial factors governing a determined phenotypic state is known as Master Regulator Analysis (MRA) [25]. In this study, we set out to apply an integrated network approach combining protein interactions and transcriptional networks to identify Master Regulators of coronavirus infections in human airway cells, in the hope of shedding more light on the virus/host interaction mechanism and also on the evolutionary origin of SARS-CoV-2.

## 2. Materials and Methods

### *Interactome model for SARS-CoV-2/human cell interaction*

We obtained a recently generated interactome of all the physical interactions between SARS-CoV-2 and human proteins based on the work by Korkin and colleagues [26]. In brief, the interactome contains both virus-virus and virus-host protein-protein interactions based on structural interaction modeling and established experimental knowledge. The resulting interactome consists of 125 proteins (31 viral proteins and 94 human host proteins) and 200 unique interactions. The interactome is available at <http://korkinlab.org/wuhan>.

### *Coexpression-based lung network*

Coexpression-based networks were generated using the *corto* algorithm with default parameters, freely available on the CRAN repository of R packages [27], using the SARS-CoV-2/human interactome proteins derived from [26] and the largest human Lung RNA-Seq dataset available from the GTEX consortium, generated from 427 patients gene expression profiles [28]. In brief, *corto* calculates a coexpression network for each protein and then removes indirect interactions using Data Processing Inequality [29], which has been shown to provide a more robust readout of single protein abundance alone [30].

### *Datasets of MERS and SARS infection*

We obtained two datasets describing transcriptome-wide effects of coronavirus infection of human cultured 2B4 bronchial epithelial cells. The first dataset (available as ArrayExpress [31] dataset E-GEOD-17400) measured SARS-CoV infection using the microarray platform Affymetrix Human Genome U133 Plus 2.0 [32], producing 9 infected samples and 9 mock samples. The dataset was normalized using RMA [33] and probeset mapping was performed using the most updated annotation from CustomCDF v24.0.0 [34]. The second dataset (available as ArrayExpress dataset E-GEOD-56677) measured gene expression upon MERS-CoV infection using Agilent-039494 SurePrint G3 Human GE v2 8x60K microarrays [35] and contained 18 infected samples and 15 mock samples.

### *Master Regulator Analysis*

Master Regulator Analyses were performed by comparing infected and mock samples in both MERS and SARS datasets separately with the *corto* algorithm [27] using default parameters and the coexpression network derived from human Lung samples. In brief, a gene-by-gene signature of viral-induced differential expression is generated, and a combined value for each coexpression network is generated by weighting every gene's likelihood in the network, providing a final Normalized Enrichment Score for each human/SARS-nCoV-2 interactome member, which is positive if the network is upregulated by the infection, and negative if it's downregulated, as in [36].

### *BLAST*

Pairwise protein identity and coverage was calculated using BLAST protein v2.6.0 [20] with BLOSUM62 matrix and default parameters. Nucleotide sequence identity and coverage were calculated using BLAST nucleotide v2.6.0 [20].

### *Phylogenetic Analysis*

201 genome sequences were obtained from GISAID and Genbank. Specifically, we obtained 8 pangolin coronavirus sequences, 6 bat coronavirus sequences, 6 SARS coronavirus sequences, 3 MERS coronavirus sequences and all 177 SARS-nCoV-2 genome sequences available on 25 February 2020. A full description of all GISAID sequences is available at Supplementary Table S1. To obtain the pangolin virus sequence PRJNA573298 we performed a deeper data mining analysis. FASTQ sequences were retrieved from the Sequence Read Archive (SRA) using SRA-toolkit v2.9.6-1 [37] using project PRJNA573298 run ids SRR10168377 and SRR10168378, which provided a DNA readout of the pangolin viral metagenome [38]. Reads were mapped over the reference 2019-nCoV genome NC\_045512.2 using Hisat2 v2.1.0 [39] with default parameters (with option --no-spliced-alignment). Environment programs were samtools v1.9 and bedtools v 2.26.0. Genome assembly of the virus genome-mapping reads was performed using Abyss v 2.1.5 with default parameters [40].

Multiple Sequence Alignment (MSA) of all 201 genomic sequences was performed using MUSCLE v3.8.31 [41] and is available in FASTA format as Supplementary File S1. MSA visualization was generated *via* Jalview v 2.11.0 [42]. Phylogenetic model generation and tree visualization was done using MEGAX v 10.1.7 [43], using the Maximum Likelihood method and Tamura-Nei model [44]. The tree structure was validated by running the analysis on 100 bootstrapped input datasets [45]. Branching points with bootstrap validation below 30% were collapsed. The tree in Figure 4 collapsed the SARS-nCoV-2 subtree, but a full tree is available at Supplementary Figure S1.

### *3D structural analysis*

The protein structure representation has been produced with Chimera v1.14 [46] based on Protein Data Bank structure id 1R42 [47].

## 3. Results

### *3.1 Human/COVID-19 interactome response to coronavirus infection*

We obtained the most up-to-date specific SARS-CoV-2/human interactome from [26]. This network, which we represent in Figure 1, highlights the complex interaction between the 31 viral proteins (coded by a surprisingly compact genome of only 30kb) and 94 human proteins, ranging from surface receptors (ACE2) to RNA-processing proteins (DDX5) to mitochondrial constituents (BCL2). In order to understand how this interactome is functionally affected by the viral infection, we obtained transcriptome wide microarray data depicting gene expression changes upon MERS-CoV [35] and SARS-CoV [32] on human bronchial-derived 2B4 cells. We therefore performed Master Regulator Analysis (MRA) [22] using a context-specific network model derived from the largest available human Lung RNA-Seq dataset from GTEx [28]. This allowed us to have a robust readout not only of the changes of the transcript of each protein-encoding human gene in the interactome, but also of its proximal functional network [30]. Since transcriptomics data of SARS-CoV-2 infection is not yet available, we reasoned that combining the SARS-CoV-2 specific interactome with transcriptome-wide datasets from two viruses belonging to the same family would provide a robust readout of the human cell response to this new virus.

The combined results of both MERS-CoV and SARS-CoV infection are depicted in Figure 1 using Stouffer integration of the Normalized Enrichment Scores (NES) of both analyses. Individual MRAs of beta-coronavirus infections are available in Figure 2 for MERS-CoV analysis (panel A) and SARS-CoV analysis (panel B) for the 8 protein networks with the highest integrated NES. Figure 3 shows

the compared MRAs of MERS-CoV and SARS-CoV, and we decided to focus on the 8 proteins that share the highest response similarity in both experiments, as they are more likely to be conserved host responses to beta-coronavirus infection. Full results for all human interactome proteins are available in Supplementary File S2, for SARS-CoV, MERS-CoV and both integrated by Stouffer integration.

Our MRA results show a strong, shared upregulation of MCL1, a positive regulator of apoptosis [48]. Apoptosis is a complex molecular cascade usually activated to eliminate unwanted cells in a living organism, for example during development, and it can be self-initiated or induced by immune system cells [49]. Also during virus attacks, induced cell death is a known mechanism to get rid of infected cells: specifically, the activation of MCL1 in coronavirus infection has been observed as an early host cell mechanism of antiviral defense [50]. It is therefore likely that the host cell is able to sense the virus, perhaps even by the direct interaction between MCL1 and the viral ORF7a (Figure 1), to start the apoptotic cascade and shut down both itself and the replicating viruses. However, apoptosis has been reported to help the spreading of the virus by increasing replication rates and virion release from the dying cell in avian coronaviruses [51].

We also report several concordantly human proteins down-regulated by beta-coronavirus infections. One of these is EEF1A1 (Elongation factor 1-alpha 1), which is involved in tRNA delivery to the ribosome, and is known to be activated upregulated upon inflammation [52]. This protein has been shown to physically interact to proteins of several viral species, specifically hepatitis delta [53] and avian reoviruses [54]; in the latter, knock-down of EEF1A1 has been shown to significantly impair the virus infection cycle. Therefore, the downregulation of EEF1A1 could be the result of an innate cell strategy to deprive SARS-CoV-2 of a key support for RNA replication.

Similar to EEF1A1, the NDUFA10 protein network is downregulated by viral infection. NDUFA10 is a member of the mitochondrial complex I [55]. This protein is known to be shut-down upon viral infection of lung A549 cells by human respiratory syncytial virus (HRSV), a paramyxovirus, also RNA-based [56]. The down-regulation of mitochondrial components in both coronaviruses and paramyxoviruses shows the conservation of RNA viruses strategies to attack the host cell, which passes through the disruption of mitochondria.

Another down-regulated network is centered around the T cell anergy-related E3 ubiquitin ligase RNF128, also called GRAIL [57], which has been reported to provide *positive* reinforcement of antiviral immune response to RNA viruses [58]. In this case, the inactivation of RNF128, which gets physically bound by two viral nonstructural proteins, NSP7 and NSP13, could be a coronavirus resistance mechanism against the cell innate defense.

MRA shows up-regulation (significant in MERS infection, not significant in SARS) of DDX5, DEAD-box polypeptide 5, a prototypical member of the RNA helicase family. DDX5 plays multifunctional roles involved in all aspects of RNA metabolism [59]. Since SARS-CoV-2, like all coronaviruses, is a single strand RNA virus, it uses both its own RNA-dependent RNA polymerase and host proteins to promote the replication of its genetic material. The up-regulation of DDX5 could be another mechanism of viral-triggered cell activation, especially since this protein has been shown to promote the infection capability of encephalitis flaviviruses [60], which are also RNA viruses.

It is important to note that all protein activities are calculated based on physiological coexpression network models, but they are based on largely non-overlapping coexpressed genes, thanks to the DPI step in the *corto* algorithm. This allows to identify complementary signatures and co-responding networks based on independent gene measurements.

### 3.2 ACE2: implications for coronavirus origin

Perhaps the most well-known human protein in relation with beta-coronavirus infection is ACE2 (Angiotensin Converting Enzyme 2), a zinc metalloprotease able to hydrolyze angiotensin I, angiotensin II, apelin-13 and dynorphin A 1-13 and 7 additional small peptides [61]. ACE2 is exploited by SARS-CoV as its cellular entrance, which is based on the initial interaction between the viral S (spike) protein and ACE2 [62]. Given the similarity between SARS-CoV and SARS-CoV-2 sequences and symptoms it has been suggested that it is the main entry target for both viruses [63]. Its down-regulation upon SARS-CoV and MERS-CoV infection could intuitively be an innate cell defense mechanism to remove this extracellular marker to reduce the capability of the virus to enter the host cell, while at the same time reducing the RNA metabolic machinery through EEF1A1 down-regulation. ACE2 is also targeted by the neuraminidase protein of influenza A virus [64]. However, the down-regulation of ACE2 has also been shown to be beneficial to viral infection, as it can trigger lung injury and faster cell-to-cell virus spreading [65]. In influenza A (cause by orthomyxoviruses, also RNA-based) high levels of ACE2 have in fact been correlated to increased resistance to the infection [64]. So, like in apoptosis induction, the down-regulation of ACE2 could be a SARS-CoV-2-induced mechanism, from which the virus gains in capability to spread faster by damaging the host lung tissue.

Our analysis so far has shown effects on human proteins interacting SARS-CoV-2 based on SARS-CoV and MERS-CoV signatures. We conjectured that proteins similarly affected by both infections (like ACE2) could have a similar role in all beta-coronavirus infections, including SARS-CoV-2. Our analysis confirms previous reports [14] that SARS-CoV and MERS-CoV are indeed highly related to SARS-CoV-2, with genome sequence identity of ~80%. However, these two human viruses are not the most likely most direct ancestors of SARS-CoV-2, as highlighted by our phylogenetic analysis (updated on 28-February-2020) based on several beta-coronaviruses (Figure 4). The evolutionary history model shows that SARS and MERS viruses are more distantly related to SARS-CoV-2 than wild beta-coronaviruses infecting bats and pangolins. Indeed, there has been an open debate whether pangolins or bats have been the most direct culprits for the zoonotic origin of this virus [66]. Our analysis shows that at least one bat beta-coronavirus sequence derived from the GISAID database (EPI\_ISL\_402131) appears to be the closest to the group of human SARS-CoV-2 genomes. However, this is based on a low bootstrap tree branching confidence (37%), and we observed a large quantity of pangolin beta-coronaviruses which are likely to be equidistant from the human causes of COVID-19.

Based on our confirmation that ACE2 is an extremely important host protein for viral infection, we set out to compare ACE2 sequences in both bats and pangolins, to detect the most similar sequence to human ACE2. We obtained representative sequences from human ACE2 (NP\_068576.1), pangolin ACE2 (*Manis javanica* XP\_017505752.1) and bat ACE2 (AGZ48803.1, from the species *Rhinolophus sinicus*, being the closest species with a full protein sequence to *Rhinolophus affinis*, where the closest bat\_CoV was retrieved). These sequences appear to be conserved in these species, having the exact same aminoacid length (805aa). The sequence identity between the human and pangolin ACE2 is 84.76%, while the identity between human and bat ACE2 is lower: 80.60% (Figure 5 A). In Figure 5 B, we provide a visualization of the human ACE2 structure, inferred via X-ray crystallography at 2.20 Å, one of the highest resolutions for this protein [67], and highlighting all the residuals identical in human and pangolins, but different between humans and bats. Finally, the Multiple Sequence Alignment (MSA) of the proteins from the three mammals reports the specific ACE2 aminoacids that differ between the species (Figure 5 C and Supplementary Table S2). The higher similarity between human and pangolin ACE2 proteins could support the hypothesis that pangolin is a better candidate for the origin of SARS-CoV-2.

#### 4. Discussion

Our analysis integrates genome-wide and transcriptome-wide complementary network information on SARS-CoV-2/human interaction to provide a map of host proteins affected by the viral infection. Given the importance of finding new targets to effectively treat this new virus, understanding the molecular effects of this virus on human proteins might be pivotal in prioritizing pharmacological strategies. One of the SARS-CoV-2-interacting protein networks most down-regulated by beta-coronaviruses in our analysis was ACE2, and in fact the administration of recombinant ACE2 to compensate this protein loss has been shown to be effective in treating acute respiratory distress syndrome, which is often caused by severe pulmonary infections [68]. This recombinant ACE2, called APN-01, has been proposed as a potential cure for SARS-CoV-2 most severe symptoms [69–71], which could be extremely helpful given the observed clinical inefficacy of common antiviral drugs, like corticosteroids [72].

Another point of clinical intervention could be to down-regulate apoptotic mechanisms induced by the virus, e.g. by our reported MCL1 activation upon both SARS-CoV and MERS-CoV infections. As we reported, this response is in agreement to previous findings discussing the involvement of MCL1 upregulation in viral replication across many RNA-based viruses [50], possibly resulting from protein synthesis impairment eliciting ER-stress-induced apoptosis via unfolded protein response (UPR) [73]. Since beta-coronavirus-induced UPR has been shown to modulate innate immunity [74], targeting aberrant monocytes/macrophages activation may be another possible intervention strategy to reduce severity of coronavirus-induced symptomatology.

We also highlighted how mitochondrial proteins interacting with SARS-CoV-2 appear to be downregulated in beta-coronavirus infections as a general viral mechanism to control host cell metabolism and ultimately steering apoptosis [75]. The goal of preventing mitochondria-induced apoptosis is shared by clinical strategies to slow neurodegenerative pathologies, mainly Alzheimer's disease, where mitochondrial dysfunctions are directly linked to neuronal death [76]. Our study highlights several other proteins that could be directly repressed or enhanced by specific drugs. For example, a mildly virus-activated protein in our network is the serine/threonine kinase MARK2 (Figure 1), which is one of the target of the drug Fostamatinib in the treatment of Rheumatoid Arthritis and Immune Thrombocytopenic Purpura [77]. Apart from drug repurposing [78], the recent advances in drug discovery by virtual screening [79] give scientists the chance to quickly develop novel molecules to specifically target not only viral proteins but also the human proteins on whose interaction SARS-CoV-2 depends for infection and replication [80].

## Figure Legends

Figure 1. Representation of the predicted SARS-CoV-2/Human interactome [26] (available for download at <http://korkinlab.org/wuhanDataset>), containing 200 unique interactions among 125 proteins (nodes). SARS-CoV-2 proteins are depicted as green circles, while human proteins are represented as squares. The color of human protein nodes reflects the integrated effect of MERS and SARS infections on the node network (see Figure 3) as Normalized Enrichment Score (NES). Network visualization was performed via Cytoscape [81].

Figure 2. Master Regulator Analysis of the 8 human proteins in the human/SARS-CoV-2 interactome and most concordantly affected by beta-coronavirus infection. The visualization was obtained through the *mrplot* function of the R CRAN package *corto*. In brief, for each analysed network, the centroid is indicated by its gene symbol (e.g. *EEF1A1*, *ACE2*). The genes in each network (generated by the *corto* package from the GTEX healthy lung RNA-Seq dataset) are shown in a barcode-like diagram showing all transcriptome genes by means of their differential expression upon viral infection, from most downregulated (left) to most upregulated (right). Positively- (red) and negatively- (blue) correlated targets are overlaid on the differential expression signature as bars of different color. Normalized Enrichment Score (NES) and p-value are also indicated. To the right, the 12 highest-likelihood network putative targets of each protein are shown, in red if upregulated, in blue if downregulated, with a pointed arrow if predicted to be activated by the centroid protein, and with a blunt arrow if predicted to be repressed. The figure shows two analyses based on the MERS infection signature (A) and on the SARS infection signature (B). Full results are available as Supplementary File S2.

Figure 3. Scatterplot depicting Normalized Enrichment Scores (NES) of Master Regulator Analysis of human proteins interacting with SARS-CoV-2. Two analyses are compared using the signatures of MERS and SARS infection on human bronchial 2B4 cells. The 8 most significant proteins by absolute sum of NES are labeled. Full results are available as Supplementary File S2.

Figure 4. Maximum Likelihood Phylogenetic tree of selected beta coronaviruses: MERS (in green), SARS (in blue), bat (in red), pangolin (in orange) and SARS-CoV-2 (in black). Full tree including all SARS-CoV-2 sequences is available as Supplementary Figure S1.

Figure 5. Cross-species analysis of the Angiotensin-converting enzyme 2 (ACE2) protein. (A) Diagram representing sequence identity between the Human, Bat and Pangolin ACE2 proteins described in this study. (B) A visualization of a human ACE2 crystal structure (resolution: 2.2 Å, PDB:1R42). The residues that are conserved across human, pangolin and bat are depicted in cornflower blue, instead, the residues that are conserved in human and pangolin but differ in bat (Supplementary Table S2) are depicted in opaque pink. (C) MSA of the Human, Pangolin and Bat ACE2, with ClustalX coloring, conservation, quality and consensus logo indicated.

Supplementary Materials: Figure S1: full phylogenetic tree of beta-coronaviruses, generated with Maximum Likelihood algorithms and 100 bootstraps. File S1: MSA of beta-coronavirus genome sequences used in this study, in aligned FASTA format. File S2: full results of Master Regulator Analysis on MERS and SARS infections of human bronchial cells, expressed as NES. Table S1: summary of all Gisaid genomic sequences used in this study. Table S2: comprehensive list of aminoacidic differences between human/pangolin ACE2 and bat ACE2.

Funding: This research was supported by CINECA, grant numbers HP10CPQJBV and HP10CC5F89; this research was also funded by the Italian Ministry of University and research.

Acknowledgments: We thank our families and colleagues for their support, specifically Luca Triboli, Lupo Giorgi, Eleonora Fornasari.

Conflicts of Interest: The authors declare no conflict of interest.

## References

1. Cui, J.; Li, F.; Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 2019, *17*, 181–192.
2. Tortorici, M.A.; Veesler, D. Structural insights into coronavirus entry. In *Advances in Virus Research*; Elsevier, 2019; Vol. 105, pp. 93–116 ISBN 978-0-12-818456-1.
3. Zeng, Z.-Q.; Chen, D.-H.; Tan, W.-P.; Qiu, S.-Y.; Xu, D.; Liang, H.-X.; Chen, M.-X.; Li, X.; Lin, Z.-S.; Liu, W.-K.; et al. Epidemiology and clinical characteristics of human coronaviruses OC43, 229E, NL63, and HKU1: a study of hospitalized children with acute respiratory tract infection in Guangzhou, China. *Eur. J. Clin. Microbiol. Infect. Dis.* 2018, *37*, 363–369.
4. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* 2020, *382*, 727–733.
5. Zhong, N.; Zheng, B.; Li, Y.; Poon, L.; Xie, Z.; Chan, K.; Li, P.; Tan, S.; Chang, Q.; Xie, J.; et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *The Lancet* 2003, *362*, 1353–1358.
6. Zaki, A.M.; van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.D.M.E.; Fouchier, R.A.M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* 2012, *367*, 1814–1820.
7. Paules, C.I.; Marston, H.D.; Fauci, A.S. Coronavirus Infections—More Than Just the Common Cold. *JAMA* 2020, *323*, 707.
8. Wan, Y.; Shang, J.; Graham, R.; Baric, R.S.; Li, F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J. Virol.* 2020, JVI.00127-20, jvi;JVI.00127-20v1.
9. Song, W.; Gui, M.; Wang, X.; Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLOS Pathog.* 2018, *14*, e1007236.
10. Park, Y.-J.; Walls, A.C.; Wang, Z.; Sauer, M.M.; Li, W.; Tortorici, M.A.; Bosch, B.-J.; DiMaio, F.; Veesler, D. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nat. Struct. Mol. Biol.* 2019, *26*, 1151–1157.
11. Wang, D.; Hu, B.; Hu, C.; Zhu, F.; Liu, X.; Zhang, J.; Wang, B.; Xiang, H.; Cheng, Z.; Xiong, Y.; et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA* 2020.
12. Wu, Z.; McGoogan, J.M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020.

13. Mahase, E. Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ* 2020, m641.
14. Ceraolo, C.; Giorgi, F.M. Genomic variance of the 2019-nCoV coronavirus. *J. Med. Virol.* 2020, jmv.25700.
15. Coutard, B.; Valle, C.; de Lamballerie, X.; Canard, B.; Seidah, N.G.; Decroly, E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* 2020, 176, 104742.
16. Li, R.; Qiao, S.; Zhang, G. Analysis of angiotensin-converting enzyme 2 (ACE2) from different species sheds some light on cross-species receptor usage of a novel coronavirus 2019-nCoV. *J. Infect.* 2020, S0163445320300906.
17. Xu, H.; Zhong, L.; Deng, J.; Peng, J.; Dan, H.; Zeng, X.; Li, T.; Chen, Q. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int. J. Oral Sci.* 2020, 12, 8.
18. Woo, P.C.Y.; Huang, Y.; Lau, S.K.P.; Yuen, K.-Y. Coronavirus Genomics and Bioinformatics Analysis. *Viruses* 2010, 2, 1804–1820.
19. Lakshmi, B.S.; Abillasha, M.L.; Kanguane, P. Emerging Technologies for Antiviral Drug Discovery. In *Global Virology III: Virology in the 21st Century*; Shapshak, P., Balaji, S., Kanguane, P., Chiappelli, F., Somboonwit, C., Menezes, L.J., Sinnott, J.T., Eds.; Springer International Publishing: Cham, 2019; pp. 59–71 ISBN 978-3-030-29021-4.
20. Caldera, M.; Buphamalai, P.; Müller, F.; Menche, J. Interactome-based approaches to human disease. *Curr. Opin. Syst. Biol.* 2017, 3, 88–94.
21. Mercatelli, D.; Scalambra, L.; Triboli, L.; Ray, F.; Giorgi, F.M. Gene regulatory network inference resources: A practical overview. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 2019, 194430.
22. Lefebvre, C.; Rajbhandari, P.; Alvarez, M.J.; Bandaru, P.; Lim, W.K.; Sato, M.; Wang, K.; Sumazin, P.; Kustagi, M.; Bisikirska, B.C.; et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* 2010, 6, 377.
23. Giorgi, F.M.; Lopez, G.; Woo, J.H.; Bisikirska, B.; Califano, A.; Bansal, M. Inferring protein modulation from gene expression data using conditional mutual information. *PLoS One* 2014, 9, e109569.
24. Holding, A.N.; Giorgi, F.M.; Donnelly, A.; Cullen, A.E.; Nagarajan, S.; Selth, L.A.; Markowitz, F. VULCAN integrates ChIP-seq with patient-derived co-expression networks to identify GRHL2 as a key co-regulator of ERα at enhancers in breast cancer. *Genome Biol.* 2019, 20, 91.
25. Tovar, H.; García-Herrera, R.; Espinal-Enríquez, J.; Hernández-Lemus, E. Transcriptional master regulator analysis in breast cancer genetic networks. *Comput. Biol. Chem.* 2015, 59, 67–77.
26. Cui, H.; Gao, Z.; Liu, M.; Lu, S.; Mo, S.; Mkandawire, W.; Narykov, O.; Srinivasan, S.; Korkin, D. *Structural genomics and interactomics of 2019 Wuhan novel coronavirus, 2019-nCoV, indicate evolutionary conserved functional regions of viral proteins*; Bioinformatics, 2020;
27. Mercatelli, D.; Lopez-Garcia, G.; Giorgi, F.M. *corto : a lightweight R package for Gene Network Inference and Master Regulator Analysis*; Bioinformatics, 2020;
28. The GTEx Consortium; Ardlie, K.G.; Deluca, D.S.; Segre, A.V.; Sullivan, T.J.; Young, T.R.; Gelfand, E.T.; Trowbridge, C.A.; Maller, J.B.; Tukiainen, T.; et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* 2015, 348, 648–660.
29. Reverter, A.; Chan, E.K.F. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 2008, 24, 2491–2497.
30. Mercatelli, D.; Ray, F.; Giorgi, F.M. Pan-Cancer and Single-Cell Modeling of Genomic Alterations Through Gene Expression. *Front. Genet.* 2019, 10, 671.

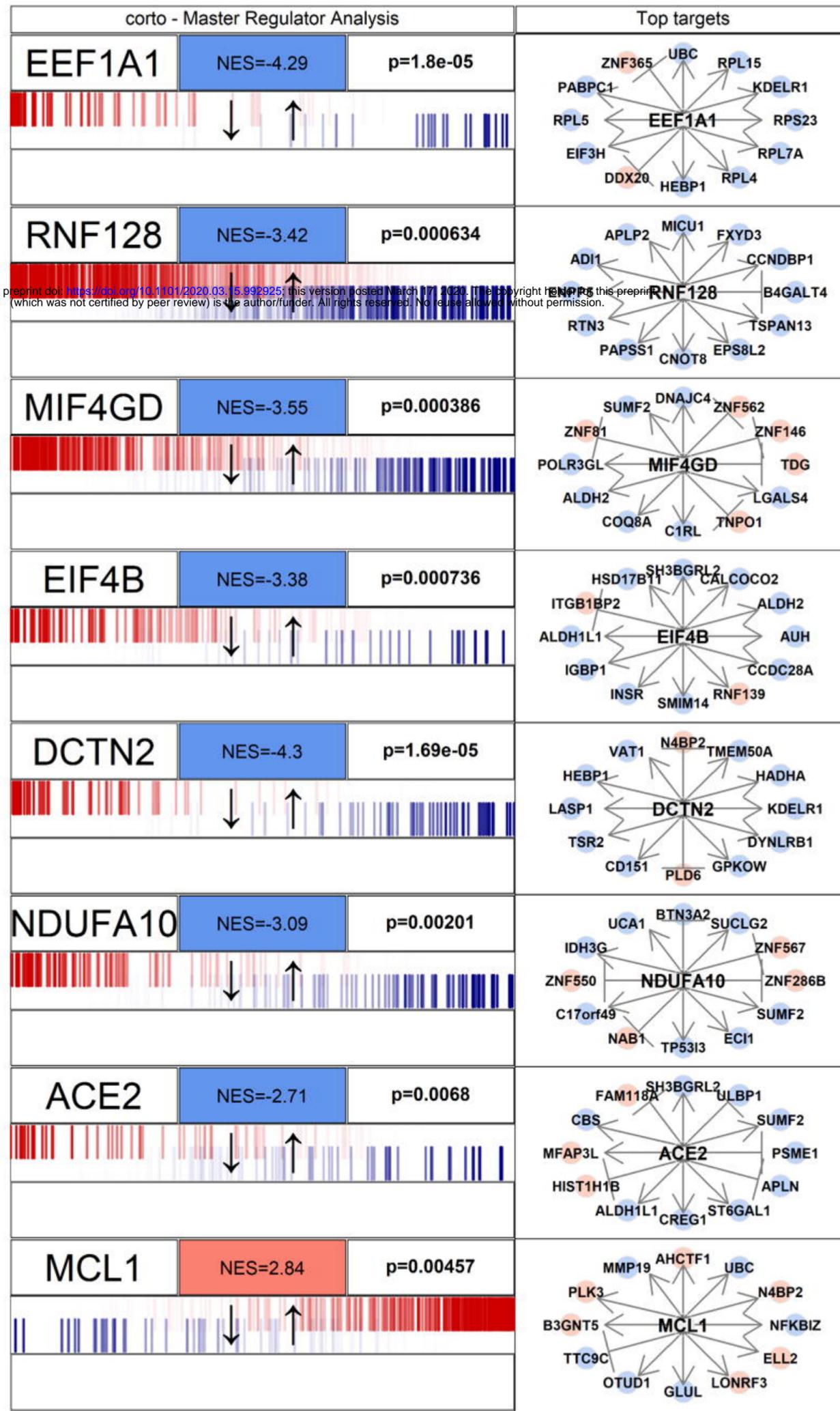
31. Kolesnikov, N.; Hastings, E.; Keays, M.; Melnichuk, O.; Tang, Y.A.; Williams, E.; Dylag, M.; Kurbatova, N.; Brandizi, M.; Burdett, T.; et al. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.* 2015, *43*, D1113–D1116.
32. Yoshikawa, T.; Hill, T.E.; Yoshikawa, N.; Popov, V.L.; Galindo, C.L.; Garner, H.R.; Peters, C.J.; Tseng, C.-T. (Kent) Dynamic Innate Immune Responses of Human Bronchial Epithelial Cells to Severe Acute Respiratory Syndrome-Associated Coronavirus Infection. *PLoS ONE* 2010, *5*, e8729.
33. Giorgi, F.M.; Bolger, A.M.; Lohse, M.; Usadel, B. Algorithm-driven artifacts in median polish summarization of microarray data. *BMC Bioinformatics* 2010, *11*, 553.
34. Dai, M. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 2005, *33*, e175–e175.
35. Selinger, C.; Tisoncik-Go, J.; Menachery, V.D.; Agnihothram, S.; Law, G.; Chang, J.; Kelly, S.M.; Sova, P.; Baric, R.S.; Katze, M.G. Cytokine systems approach demonstrates differences in innate and pro-inflammatory host responses between genetically distinct MERS-CoV isolates. *BMC Genomics* 2014, *15*, 1161.
36. Alvarez, M.J.; Shen, Y.; Giorgi, F.M.; Lachmann, A.; Ding, B.B.; Ye, B.H.; Califano, A. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* 2016, *48*, 838–847.
37. Kodama, Y.; Shumway, M.; Leinonen, R.; on behalf of the International Nucleotide Sequence Database Collaboration The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.* 2012, *40*, D54–D56.
38. Liu, P.; Chen, W.; Chen, J.-P. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (*Manis javanica*). *Viruses* 2019, *11*, 979.
39. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 2015, *12*, 357–360.
40. Jackman, S.D.; Vandervalk, B.P.; Mohamadi, H.; Chu, J.; Yeo, S.; Hammond, S.A.; Jahesh, G.; Khan, H.; Coombe, L.; Warren, R.L.; et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* 2017, *27*, 768–777.
41. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004, *32*, 1792–1797.
42. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009, *25*, 1189–1191.
43. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 2018, *35*, 1547–1549.
44. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 1993.
45. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985, *39*, 783–791.
46. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004, *25*, 1605–1612.
47. Burley, S.K.; Berman, H.M.; Kleywegt, G.J.; Markley, J.L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. In *Protein Crystallography*; Wlodawer, A., Dauter, Z., Jaskolski, M., Eds.; Springer New York: New York, NY, 2017; Vol. 1607, pp. 627–641 ISBN 978-1-4939-6998-2.

48. Kozopas, K.M.; Yang, T.; Buchan, H.L.; Zhou, P.; Craig, R.W. MCL1, a gene expressed in programmed myeloid cell differentiation, has sequence similarity to BCL2. *Proc. Natl. Acad. Sci. U. S. A.* 1993, *90*, 3516–3520.
49. Ren, J.; Shi, M.; Liu, R.; Yang, Q.-H.; Johnson, T.; Skarnes, W.C.; Du, C. The Birc6 (Bruce) gene regulates p53 and the mitochondrial pathway of apoptosis and is essential for mouse embryonic development. *Proc. Natl. Acad. Sci.* 2005, *102*, 565–570.
50. Zhong, Y.; Liao, Y.; Fang, S.; Tam, J.P.; Liu, D.X. Up-Regulation of Mcl-1 and Bak by Coronavirus Infection of Human, Avian and Animal Cells Modulates Apoptosis and Viral Replication. *PLoS ONE* 2012, *7*, e30191.
51. Liu, C.; Xu, H.Y.; Liu, D.X. Induction of Caspase-Dependent Apoptosis in Cultured Cells by the Avian Coronavirus Infectious Bronchitis Virus. *J. Virol.* 2001, *75*, 6402–6409.
52. Maruyama, T.; Nara, K.; Yoshikawa, H.; Suzuki, N. Txk, a member of the non-receptor tyrosine kinase of the Tec family, forms a complex with poly(ADP-ribose) polymerase 1 and elongation factor 1 $\alpha$  and regulates interferon- $\gamma$  gene transcription in Th1 cells: Txk, PARP1 and EF-1 $\alpha$  make a Th1 cell transcription factor. *Clin. Exp. Immunol.* 2006, *147*, 164–175.
53. Sikora, D.; Greco-Stewart, V.S.; Miron, P.; Pelchat, M. The hepatitis delta virus RNA genome interacts with eEF1A1, p54nrb, hnRNP-L, GAPDH and ASF/SF2. *Virology* 2009, *390*, 71–78.
54. Zhang, Z.; Lin, W.; Li, X.; Cao, H.; Wang, Y.; Zheng, S.J. Critical role of eukaryotic elongation factor 1 alpha 1 (EEF1A1) in avian reovirus sigma-C-induced apoptosis and inhibition of viral growth. *Arch. Virol.* 2015, *160*, 1449–1461.
55. Loeffen, J.L.C.M.; Triepels, R.H.; van den Heuvel, L.P.; Schuelke, M.; Buskens, C.A.F.; Smeets, R.J.P.; Trijbels, J.M.F.; Smeitink, J.A.M. cDNA of Eight Nuclear Encoded Subunits of NADH:Ubiquinone Oxidoreductase: Human Complex I cDNA Characterization Completed. *Biochem. Biophys. Res. Commun.* 1998, *253*, 415–422.
56. C. Munday, D.; Hiscox, J.A.; Barr, J.N. Quantitative proteomic analysis of A549 cells infected with human respiratory syncytial virus subgroup B using SILAC coupled to LC-MS/MS. *PROTEOMICS* 2010, *10*, 4320–4334.
57. Anandasabapathy, N.; Ford, G.S.; Bloom, D.; Holness, C.; Paragas, V.; Seroogy, C.; Skrenta, H.; Hollenhorst, M.; Fathman, C.G.; Soares, L. GRAIL. *Immunity* 2003, *18*, 535–547.
58. Song, G.; Liu, B.; Li, Z.; Wu, H.; Wang, P.; Zhao, K.; Jiang, G.; Zhang, L.; Gao, C. E3 ubiquitin ligase RNF128 promotes innate antiviral immunity through K63-linked ubiquitination of TBK1. *Nat. Immunol.* 2016, *17*, 1342–1351.
59. Jalal, C.; Uhlmann-Schiffler, H.; Stahl, H. Redundant role of DEAD box proteins p68 (Ddx5) and p72/p82 (Ddx17) in ribosome biogenesis and cell proliferation. *Nucleic Acids Res.* 2007, *35*, 3590–3601.
60. Li, C.; Ge, L.; Li, P.; Wang, Y.; Sun, M.; Huang, L.; Ishag, H.; Di, D.; Shen, Z.; Fan, W.; et al. The DEAD-box RNA helicase DDX5 acts as a positive regulator of Japanese encephalitis virus replication by binding to viral 3' UTR. *Antiviral Res.* 2013, *100*, 487–499.
61. Vickers, C.; Hales, P.; Kaushik, V.; Dick, L.; Gavin, J.; Tang, J.; Godbout, K.; Parsons, T.; Baronas, E.; Hsieh, F.; et al. Hydrolysis of Biological Peptides by Human Angiotensin-converting Enzyme-related Carboxypeptidase. *J. Biol. Chem.* 2002, *277*, 14838–14843.
62. Hofmann, H.; Pöhlmann, S. Cellular entry of the SARS coronavirus. *Trends Microbiol.* 2004, *12*, 466–472.
63. Yan, R.; Zhang, Y.; Guo, Y.; Xia, L.; Zhou, Q. *Structural basis for the recognition of the 2019-nCoV by human ACE2*; *Biochemistry*, 2020;

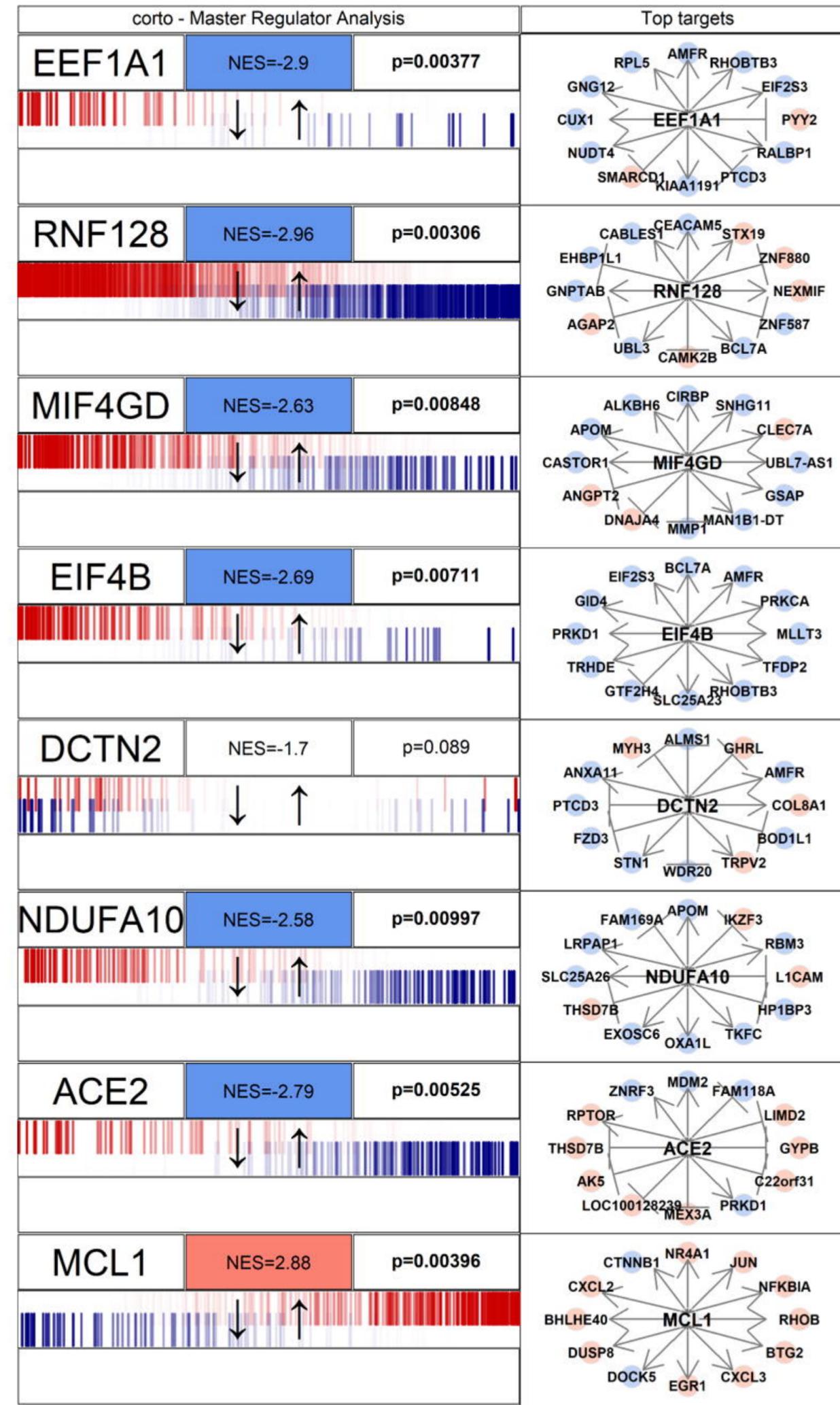
64. Liu, X.; Yang, N.; Tang, J.; Liu, S.; Luo, D.; Duan, Q.; Wang, X. Downregulation of angiotensin-converting enzyme 2 by the neuraminidase protein of influenza A (H1N1) virus. *Virus Res.* 2014, *185*, 64–71.
65. Glowacka, I.; Bertram, S.; Herzog, P.; Pfefferle, S.; Steffen, I.; Muench, M.O.; Simmons, G.; Hofmann, H.; Kuri, T.; Weber, F.; et al. Differential Downregulation of ACE2 by the Spike Proteins of Severe Acute Respiratory Syndrome Coronavirus and Human Coronavirus NL63. *J. Virol.* 2010, *84*, 1198–1205.
66. Liu, P.; Jiang, J.-Z.; Wan, X.-F.; Hua, Y.; Wang, X.; Hou, F.; Chen, J.; Zou, J.; Chen, J. *Are pangolins the intermediate host of the 2019 novel coronavirus (2019-nCoV) ?*; Pathology, 2020;
67. Towler, P.; Staker, B.; Prasad, S.G.; Menon, S.; Tang, J.; Parsons, T.; Ryan, D.; Fisher, M.; Williams, D.; Dales, N.A.; et al. ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis. *J. Biol. Chem.* 2004, *279*, 17996–18007.
68. Zhang, H.; Baker, A. Recombinant human ACE2: acing out angiotensin II in ARDS therapy. *Crit. Care* 2017, *21*, 305, s13054-017-1882-z.
69. Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020.
70. Could GSK reject benefit COVID-19 patients? Available online: <https://www.bioworld.com/articles/433328-could-gsk-reject-benefit-covid-19-patients> (accessed on Feb 28, 2020).
71. APEIRON's respiratory drug product to start pilot clinical trial to treat coronavirus disease COVID-19 in China | Proteins and Peptides | News Channels Available online: <https://pipelinereview.com/index.php/2020022673884/Proteins-and-Peptides/APEIRONs-respiratory-drug-product-to-start-pilot-clinical-trial-to-treat-coronavirus-disease-COVID-19-in-China.html> (accessed on Feb 28, 2020).
72. Russell, C.D.; Millar, J.E.; Baillie, J.K. Clinical evidence does not support corticosteroid treatment for 2019-nCoV lung injury. *The Lancet* 2020, S0140673620303172.
73. Fung, T.S.; Liu, D.X. Coronavirus infection, ER stress, apoptosis and innate immunity. *Front. Microbiol.* 2014, *5*.
74. Shi, C.-S.; Nabar, N.R.; Huang, N.-N.; Kehrl, J.H. SARS-Coronavirus Open Reading Frame-8b triggers intracellular stress pathways and activates NLRP3 inflammasomes. *Cell Death Discov.* 2019, *5*, 101.
75. Boya, P.; Pauleau, A.-L.; Poncet, D.; Gonzalez-Polo, R.-A.; Zamzami, N.; Kroemer, G. Viral proteins targeting mitochondria: controlling cell death. *Biochim. Biophys. Acta BBA - Bioenerg.* 2004, *1659*, 178–189.
76. Rao, V.K.; Carlson, E.A.; Yan, S.S. Mitochondrial permeability transition pore is a potential drug target for neurodegeneration. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* 2014, *1842*, 1267–1272.
77. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018, *46*, D1074–D1082.
78. Zhou, Y.; Hou, Y.; Shen, J.; Huang, Y.; Martin, W.; Cheng, F. *Network-based Drug Repurposing for Human Coronavirus*; Infectious Diseases (except HIV/AIDS), 2020;
79. Leelananda, S.P.; Lindert, S. Computational methods in drug discovery. *Beilstein J. Org. Chem.* 2016, *12*, 2694–2718.
80. Li, G.; De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* 2020, d41573-020-00016-0.

81. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 2003, *13*, 2498–2504.

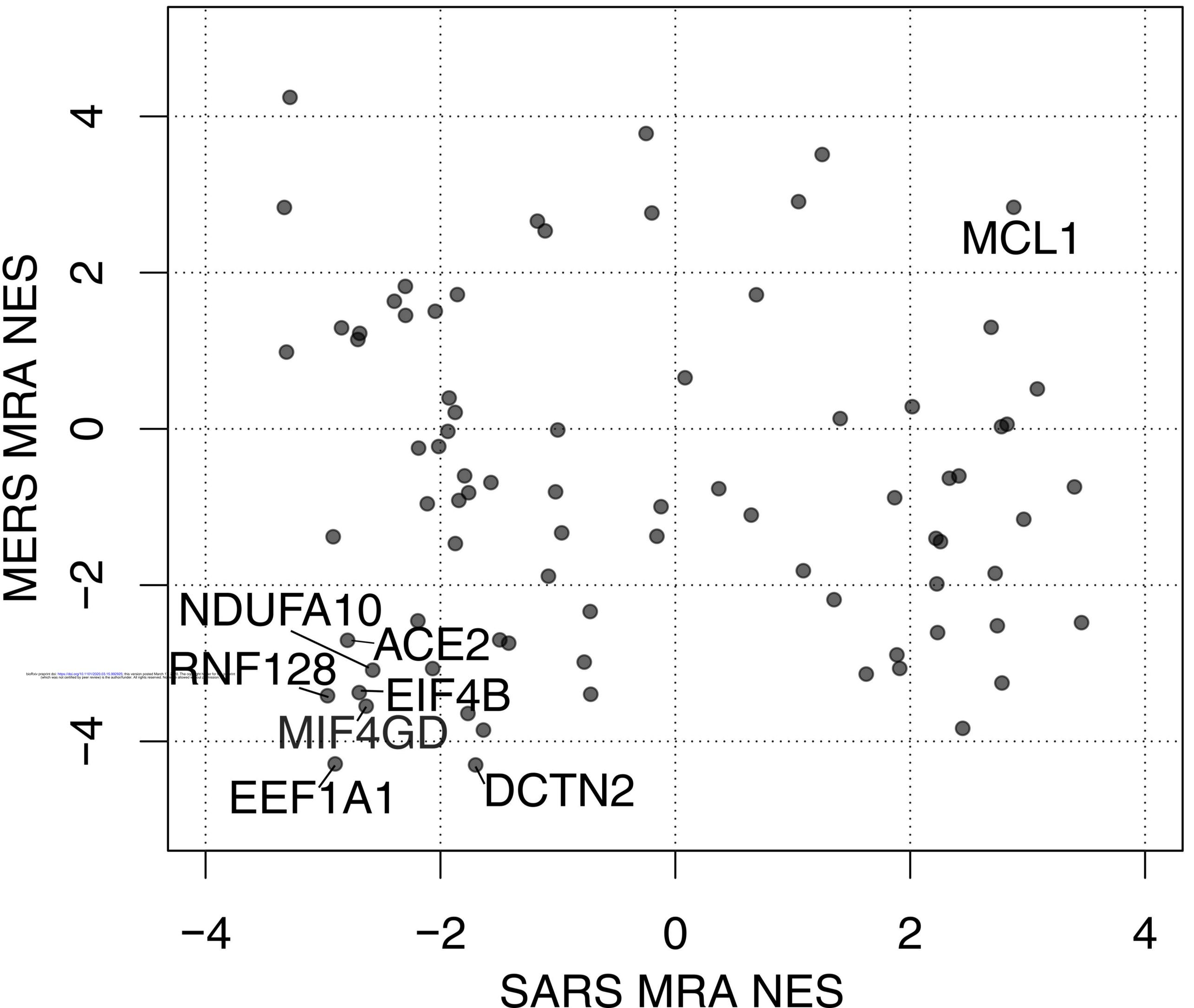


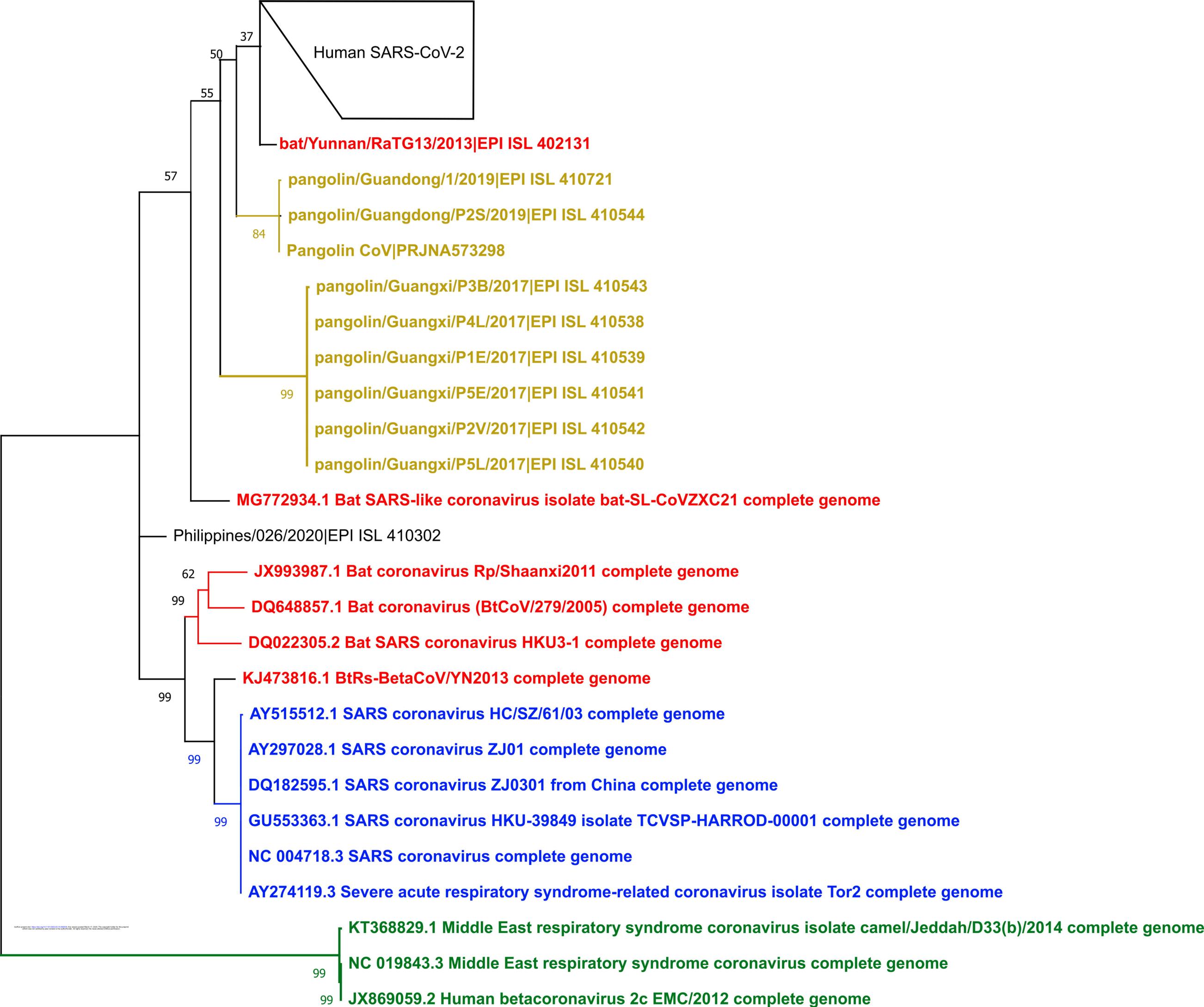
**A**

bioRxiv preprint doi: <https://doi.org/10.1101/2020.03.15.992925>; this version posted March 17, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

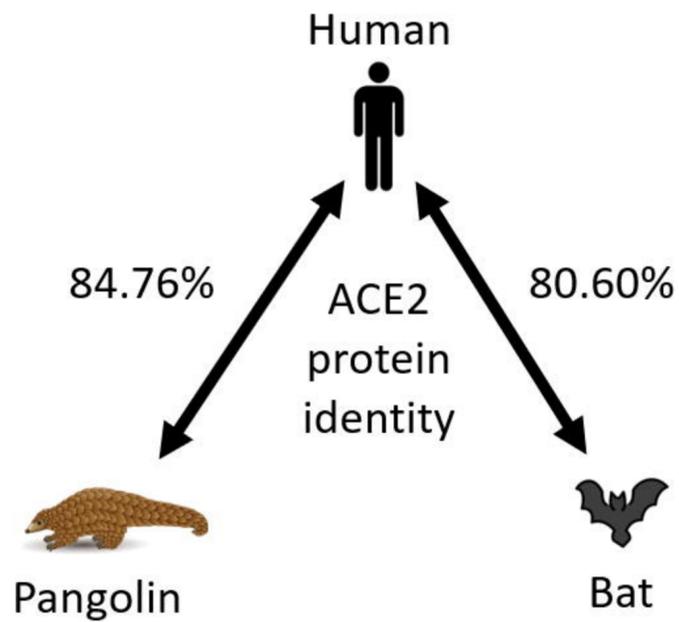
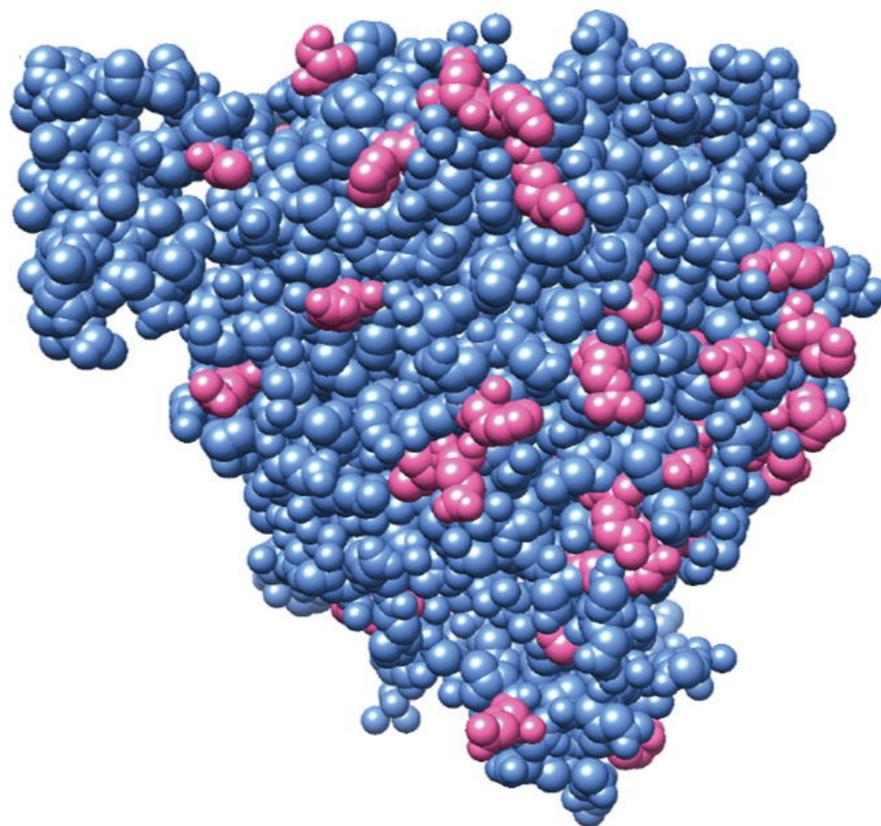
**B**

# SARS-CoV-2 Interactome Master Regulator Analysis





0.20

**A****B****C**