Article

Understanding the SARS-CoV-2-Human Liver Interactome Using a Comprehensive Analysis of the Individual Virus-Host Interactions.

[Giovanni Colonna](#) *

Article

Understanding the SARS-CoV-2-Human Liver Interactome Using a Comprehensive Analysis of the Individual Virus-Host Interactions

Giovanni Colonna

Unit of Medical Informatics—AOU-L, Vanvitelli, University of Campania, Naples, Italy; giovanni.colonna@unicampania.it

Abstract: Many metabolic processes at the molecular level support both viral attack strategies and human defenses during Covid-19. This knowledge is of vital importance in the design of antiviral drugs. In this study, we extracted 18 articles (2021-2023) from PubMed reporting the discovery of hub-nodes specific for the liver during covid-19, identifying 142 hub-nodes. They are highly connected proteins from which to get deep functional information on viral strategies when used as functional seeds. Therefore, we evaluated the functional and structural significance of each of them to endorse their reliable use as seeds. After filtering, the remaining 111 hubs were used to get by STRING an enriched interactome of 1111 nodes (13,494 interactions). It shows the viral strategy in the liver is to attack the entire cytoplasmic translational system, including ribosomes, to take control of protein biosynthesis. We used the SARS2-Human-Proteome Interaction Database (33,791 interactions), designed by us with BioGRID data to implement a reverse-engineering process that identified human proteins actively interacting with viral proteins. The results show 57% of human liver proteins directly involved in Covid, a strong impairment of the ribosome and spliceosome, an antiviral defense mechanism against cellular stress of the p53 system, and, surprisingly, a viral capacity for multiple protein attacks against single human proteins that reveal underlying evolutionary-topological molecular mechanisms. Viral behavior over time suggests different molecular strategies for different organs.

Keywords: COVID-19; Covid molecular mechanisms; SARS-CoV-2; Liver interactome; Ribosome; Liver proteome during COVID-19; viral strategy

1. Introduction

COVID-19 exhibits many characteristics of a systemic disease. Despite the limited understanding of the molecular mechanisms facilitating the virus's dissemination to distant tissues and organs, many studies are dedicated to elucidating the potential pathophysiological mechanisms associated with organ-specific infection.

A recent review [1] describes the pathological effects of the SARS-CoV-2 virus on the human liver, focusing on hepatic manifestations of COVID-19. The authors also discuss the potential pathophysiological mechanisms, as well as the diagnosis and management approaches. They report the effects described as moderate in healthy patients, describing most hepatic symptoms associated with COVID-19 as mild and self-limiting and of favourable management. Conversely, the outcomes seen in elderly, obese patients and those with previous liver disease are serious and not so manageable. What the authors complain about is the lack of a single definition of liver damage, and the lack of studies describing in more detail the cellular damage and the molecular mechanisms that generate it. According to other authors [2], coronavirus-2 RNA is detectable in liver biopsies of patients with severe acute respiratory syndrome. These authors also noted that on a histological analysis of the liver tissue sections showed many nonspecific and purely descriptive findings. Whereas viral RNA suggested viral particles that had spread from other tissues and organs. [3–5]. This type of observations has raised many considerations about the spread of the virus in organs and tissues, in particular because of complications of long-COVID-19 [6]. Although researchers have hypothesized [7] and discussed [8] general mechanisms that could inhibit programmed cell death,

the actual molecular mechanisms adopted by this virus are still unknown, and no liver-specific data is available.

The computational approach is one of the most used approaches in these cases. We can evaluate the changes in gene expression under the effect of COVID-19. We can calculate an enriched metabolic network model in which to identify high ranked genes. These are genes whose decoded products (proteoforms) show high connections with network nodes and coordinate many important metabolic processes. Through these genes, we can describe possible functions making metabolic hypotheses, aided by GO and KEGG analyses. Many researchers focused on the impact of the disease on evaluating the regeneration of liver epithelial cells [9,10], to create a model of acute liver failure and identify the possible therapeutic effects of inhibitors of some high ranked genes, such as hub genes [11]. Knowledge of hub genes, therefore, represents an indispensable crucial point for characterizing the molecular aspects of a disease, such as cellular stress due to COVID-19 but also for the design of specific drugs. Although hubs play a crucial role, there is disagreement in identifying, characterizing, and classifying these types of nodes [12].

Some of our searches on PubMed performed using terms such as "COVID-19 AND liver hub genes" or similar terms, extracted eighteen studies, of which we selected eleven (2021-2023) [13–23] because they reported hub genes linked to COVID-19 (see also in the Results). These studies aim to determine regulatory processes from datasets based on microarray or transcriptome sequencing technology. These approaches have two significant limitations. First, when studying homogenized tissue samples or disaggregated cells, the researchers lose spatial information on gene expression. And second, most of the resulting models are static and probabilistic, thus lacking any space-time reference. Therefore, they present a strong limitation in investigating a dynamic process, such as the progression of COVID-19. We should remember that studying the viral strategy and the host's innate response is best done during the initial moments of the disease (the first 3-5 days). Later, the host phenotype displays significant interferences that mask viral action and occasionally assume dominance. During the enrichment phase, we can also extract many hub genes that coordinate normal metabolic activities of the cell, regardless of COVID-19. We should identify and exclude from disease-dependent hubs those genes focused on managing fundamental cellular activities, necessary for both actors. Without delving into the reasons behind their diversity, we have still included these genes in our analysis to identify and remove them.

A network typically includes many small subgraphs; these subgraphs are under the control of several hubs, so there is no single place from which one can get a complete picture of the general and specific biological purposes of the network. Therefore, biological network analysis cannot always provide explicit support to get information about the internal parts of the network from peripheral nodes. Networks are so heterogeneous that an approach deemed useful in one condition, or in one metabolic context, may not be effective elsewhere, or at another time [24–26]. We could effectively analyze the topological differences and congested links in all selected networks from five different perspectives: data source control, topological analysis, network characteristics, validation, and prediction. But this takes a long time, and the structural/functional validation of the data is often difficult to verify. Here, we apply a biological reverse engineering protocol that involves deriving a model of the biological relationships established between the nodes implementing the networks, with no a priori knowledge of their computational protocols [27,28].

Our aim is to identify patterns of non-random connectivity or a viral organizational strategy that often remains hidden when analyzing the many molecular details of biological graphs. Many authors, through hubs, attempt to decompose biological networks into a set of network motifs with characteristic functions, often drawing biased conclusions from low-resolution data [29].

A major issue for reverse-engineered model training networks is the significant time and effort required to get and quantify spatial gene expression data [30,31]. To better understand COVID-19, we need a better and more systematic understanding of the complex regulatory networks that govern disease progression. Knowledge that is still limited. One crucial point in increasing the reliability of the calculated network models is the experimental origin of the data. Since these are one-to-one interactions between proteins, on a physical or functional basis, we should have clear experimental

confirmation, without which we do nothing but contribute to increasing the level of uncertainty of our network models.

This point is significant. In fact, relationships between lower-level processes and higher-level systems capacities are "degenerate" because of the many-to-many relationships [30,32]. The concept of degeneracy in biological systems is indeed intricate. When discussing the relationships between lower-level processes and higher-level system capacities, degeneracy refers to the situation where distinct processes (or mechanisms) within a system can perform similar functions or roles. This means that there is not a one-to-one correspondence between processes and functions; instead, multiple processes can contribute to or perform the same function, therefore the system is degenerate. In biological complexity and metabolic systems, this can be challenging. When many elements can serve the same function, understanding the underlying principles governing how these components work together becomes complex, especially when these relationships are dynamic and nonlinear. The many-to-many relationships between distinct elements and functions make it difficult to pinpoint exact cause-effect relationships or predict the system's behaviour solely based on the behaviour of individual elements. Biological complexities add layers of intricacy to understanding biological systems and their functions [33,34]. This also applies to mechanistic explanations, since they involve resorting to operations that are at a lower level to explain phenomena that are at the level of the whole mechanism [35]. This suggests that, when computing a network, functional seeds, and enrichment can discover and remove hub nodes that have non-experimental origins or uncertain functions. We apply reverse engineering based on the direct validation of biological messages exchanged between two nodes (which we call network inference). Inference comprises analyzing the experimental source of biological information by validating it with external tools. Biological networks are a key feature by which simple interactions can combine to produce complex results. We consider that biological information is transmitted through the one-to-one interactions between nodes, which are analytically represented by a network (or graph). It is often overlooked that this aspect causes the certification of the interaction as certain and reliable (refer to Appendix A for more information). Therefore, network analysis allows us to build predictive models to understand how variations in molecular interactions can influence the behaviour of a biological system. Ultimately, the reverse engineering approach offers an avenue for understanding the intricate details of COVID-19 in the liver, and may lead to useful results in understanding the molecular aspects of this disease and designing and developing therapies.

2. Materials and Methods

2.1. BioGRID

[43] is the source of experimental interactions of SARS-CoV-2 [as of July 2023]. https://thebiogrid.org/search.php?search=SARS-CoV-2*&organism=2697049

BioGRID is a general biological repository for interaction datasets. It is a curated biological database of protein-protein interactions, genetic interactions, chemical interactions, and post-translational modifications. It also collects all the experimentally proven data on the interactions between the 31 SARS-CoV-2 proteins and the human proteome. The quantitative SAINT analysis [51] was used to identify SARS-CoV-2 viral-host proximity interactions in human or model system cells [11–17] and those with a Bayesian FDR =< 0.01 were high confidence. Scores are the sum of peptide counts from four mass spec runs with a higher score indicating a higher degree of connectivity between proteins. This statistical model assigns the number of peptide identifications for each interactor to a probability distribution, which is then used to estimate the likelihood of a true interaction. In this way, the model facilitates the identification of high confidence interactomes.

2.2. STRING

STRING [44,45] [https://string-db.org/] is a proteomic database focusing on the networks and interactions of proteins in an array of species. The curated interactions are direct (physical) and indirect (functional) associations. The interactions came from 6 different sources (genomic context,

high-throughput experiments, co-expression, previous knowledge, etc.). In this paper, we established the PPI network according to the Version:11.5 of the STRING database. We constructed PPI networks by mapping proteins to the STRING database with a confidence score of 0.900 with the information of all six sources (see also note in Supplements).

STRING maps several databases onto its proteins. Therefore, this feature allows to retrieve not only the functional enrichment for any set of input proteins but also a series of specialized information on the interactome characteristics and properties. Through **InterPro**, STRING provides a functional analysis of interactome proteins by classifying them into families and predicting functional domains, i.e., those structural domains devolved into interaction to generate a function. **SMART** (Simple Modular Architecture Research Tool) allows the identification and the analysis of genetically mobile domain architectures. Through **Annotated Keyword (UniProt)**, STRING enables access to more specific information about proteins of a network, in our case to define several interaction scenarios from the liver interactome. **REACTOME** is an open-source, open access, curated and peer-reviewed pathway database mapped by STRING onto its proteins. Its goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic and clinical research, genome analysis, modeling, systems biology and education. The core unit of the Reactome data model is the reaction. Entities (nucleic acids, proteins, complexes, vaccines, anti-cancer therapeutics and small molecules) participating in reactions form a network of biological interactions and are grouped into pathways. Reactome is designed to give the user a map of known biological processes and pathways that is at the interface of its network from which the user can extract detailed information on components and their relations.

Cluster Analysis - STRING also provides the most reliable clusters in terms of compactness, metabolic functionality, and p-value, calculated on the network data (individually characterized by an acronym, CL.xxx). For the cluster analysis, it is used the K-Means Clustering method (49). K-Means Clustering is an Unsupervised Learning algorithm (centroid-based clustering algorithm) used by STRING to group the protein dataset into different functional clusters. Centroid-based algorithms are efficient, effective, simple and sensitive to initial conditions and outliers. This makes it useful in handling networks.

2.3. Protein enrichment

It is to some extent based on prior knowledge, and the statistical enrichment of the annotated features may not be an intrinsic property of the input. To get an enrichment test from STRING that is statistically valid, we must insert the entire set of enriched proteins into STRING ensuring that 'first shell' and 'second shell' are both set to 'none'. To confirm the procedure's correctness, we also checked the STRING notes to the network for a specific notice that disappears when done correctly. By adding new interaction partners to the network, we can extend the interaction neighborhood according to the required confidence score. We used 0.9 as a confidence score.

2.4. CYTOSCAPE and Network topology analysis

Cytoscape [46,47] through Network Analyzer was used to analyze the topological parameters of networks. Using Cytoscape software, we visualized and analyzed PPI networks, which offer diverse plugins for multiple analyses. Cytoscape represents PPI networks as graphs with nodes illustrating proteins and edges depicting associated interactions. We examined network architecture for topological parameters such as clustering coefficient, centralization, density, network diameter, and so on. Our analysis included undirected edges for every network. We termed the number of connected neighbours of a node in a network as the degree of a node. P(k) is used to describe the distribution of node degrees, which counts the number of nodes with degree k where k=0, 1, 2, … We calculated the power law of distribution of node degrees, which is one of the most crucial network topological characteristics. The coefficient R-Squared value (R^2), also known as the coefficient of determination, gives the proportion of variability in the dataset. We also examined other network parameters, including the distribution of various topological features. We did calculation of Hub and Bottleneck nodes based on relevant topological parameters. By examining the PPI network, we found

the top 7 hub nodes. These nodes had significantly higher degree values than the others and were primarily in two central modules that were closely connected and compact.

2.5. CentiScaPe

Centralities for undirected, directed, and weighted networks. CentiScaPe [48] computes specific centrality parameters describing the network topology. These parameters facilitate users in locating the most important nodes within a complex network. The computation of the plugin produces both numerical and graphical results, facilitating the identification of key nodes even in extensive networks. Integrating network topological quantification with other numerical node attributes can cause relevant node identification and functional classification.

2.6. GO and KEGG pathway analyses.

To better research and show the biological function of interacting proteins, we performed GO analysis, which included biological process (BP), cellular component (CC), molecular function (MF) and many other evaluations using the specific tools present in STRING. All functions shown by STRING are significant, having a p value always of <0.05.

2.7. SARS2-Human Proteome Interaction Database (SHPID)

We have collected in a single database all the files made available online by BioGRID, containing all the curated physical interactions of the 31 SARS-CoV-2 proteins gained through experiments in human cellular systems with viral baits, followed by purification and characterization with mass spectrometry. These Data are available as a zip file containing multiple zip-files (32 zip-files) each comprising Interactions and Post-Translational Modifications for each single SARS-CoV-2 protein for 33,823 interactions (as June 2023). The database therefore contains the set of all possible real interactions existing between the SARS-CoV-2 proteome with all the proteins of the human proteome. We highlight that not all interactions are real, but some could derive from artifacts of the method, such as non-biological interactions, only because of the random encounter between proteins in the system used. An encounter that would never have happened in the reality of an infection. However, the interactions derive from BioGRID where all, even those with the lowest score, have a significant statistic with an FDR =< 0.01. This allows us to identify as many significant comparisons as possible while maintaining a low false positive rate, i.e., the probability of a false positive is less than 1%, so only 338 interactions among all are truly null.

This database is the comprehensive repository of all interactions acknowledged **biologically possible** between the virus and its human host. The database also contains interactions between individual viral proteins, where known. As part of database search actions, you can ask who interacts with whom, with queries that use single human or viral proteins. The search can include multiple sets of proteins.

2.8. Comparison between GO pairs in enriched networks.

In modeled networks, STRING uses two parameters to analytically define the enriched biological terms. Strength is the measure of how large an enrichment is, expressed as Log10 [Log10 (observed/expected)], while False Discovery Rate (fdr) is the measure of the statistical significance of an enrichment given as a p-value after the Beniamini-procedure Hochberq. The higher the Strength value, the greater the biological effect because of genetic enrichment, indicating increased gene expression, which suggests a higher likelihood of the event occurring. Since STRING characterizes biological functions as pairs in which strength and fdr often show very different numerical values from each other, we use the product P [P = strength x -log10 p-value] to get a quantitative evaluation. When "Strength" has a very high value and p has a slight value, this product is enhanced (the most favorable situation for evaluating an effect is represented by the extremes of their numerical values, very high and slight, respectively). This facilitates us to compare and evaluate different pairs. Two pairs, one characterized by S = 0.35 and fdr = 1.0e-11, and another characterized by S = 1.9 and fdr =

1.0e-6, could lead one to think that the first is more significant. If we analyze the P value, we have 3.85 and 11.4, respectively. This tells us that the increase in gene expression in the second case is prevalent. The higher the value of the product, the more reliable the result of one pair will be over another. We consider that strength = 1 means a 10-fold genetic enrichment. However, it is important to remember that all fdr values reported by STRING in its biological functionality characterizations (GO, KEGG, etc.) are always significant and never greater than 0.05.

2.9. Highlighting the nodes of a STRING network involved in the same biological process (GO).

STRING makes visible all the nodes involved in the same biological process evidenced through its mapped databases onto the proteins (GO, KEGG, REACTOME, and so on) by activating the process itself with a click of the cursor on the process line. Activation means that all nodes involved in the same metabolic process stain similarly. Nodes involved in multiple processes are colored multiple times. This tool is very useful when one wants to analyze the involvement of multiple nodes in many metabolic processes visually, distinguishing the effect of different processes between nodes and identifying which nodes represent the crossing points. If individual nodes do not show any coloration under the effect of clicking, this identifies certain components of a path, or group, that a specific activated process does not influence. The relationships that determine the coloring of the nodes depend on the knowledge base that STRING organizes for a specific network by extracting data and information from the scientific literature in PubMed.

3. Results

3.1. Hub data of human liver during COVID-19

As mentioned in the Introduction, we carefully selected 11 projects [13–23] out of the 18 projects identified in the scientific literature between 2021 and 2023. These papers deal with the characterization of hepatic metabolic processes that are viral targets in patients affected by COVID-19. The distinguishing feature of these projects is the utilization of different techniques to conduct bioinformatic analyses on profiled patient genes. In particular, the authors studied the hub genes that coordinated the metabolic activities of the human liver during COVID-19 infection. They have considered them as potential drug targets for this liver pathology. Owing to their high significant rank, HUB nodes can also serve as functional seeds to extract related functions from the human proteome. By appropriately enriching the nodes that express these functions, it is possible to broaden the functional spectrum of action of the virus, accessing the mechanisms used by SARS-CoV-2 to manipulate human proteins and metabolic processes, as well as information on the molecular strategy adopted. The surprising discovery is that the hub nodes highlighted by these projects are too numerous and different from each other (**Table 1)**. Since they concern the same disease and the same virus, we should have a set of similar hub genes that control the viral strategy by inducing dysregulations in metabolic processes, but we could also come across hub nodes that coordinate normal metabolic activities (housekeeping activities).

Table 1. HUB genes found in the liver by different scientific projects during COVID-19 (2021-2023).

Article title	HUB GENES
Demonstration of the impact of COVID-19 on metabolic associated fatty liver disease by bioinformatics and system biology approach [2023].	SERPINE1, IL1RN, THBS1, TNFAIP6, GADD45B, TNFRSF12A, PLA2G7, PTGES, PTX3, and GADD45G.
Comprehensive DNA methylation profiling of COVID-19 and hepatocellular carcinoma to identify common pathogenesis and potential therapeutic targets [2023].	MYLK2, FAM83D, STC2, CCDC112, EPHX4, and **MMP1.**
Exploration and verification of COVID-19-related hub genes in liver physiological and pathological regeneration. [2023].	ASPM, **BUB1B**, CDC20, CENPF, CEP55, KIF11, KIF4, NCAPG, NUF2, NUSAP1, PBK, PTTG1, **RRM2**, TPX2, UBE2C.
Systems biology approach reveals a common molecular basis for COVID-19 and non-alcoholic fatty liver disease [NAFLD] [2022].	**IL6**, **IL1B**, PTGS2, **JUN**, **FOS**, ATF3, **SOCS3**, CSF3, NFKB2, and HBEGF.

To investigate the internal association between SARS-CoV-2 infections and cancer through bioinformatics [2022].	MMP9, **FOS**, COL1A2, COL2A1, DKK3, IHH, CYP3A4, PPARGC1A, MMP11, and APOD.
Target and drug predictions for SARS-CoV-2 infection in hepatocellular carcinoma patients [2022].	Up regulated, *PDGFRB, MMP14, VWF, CD34, NES, MCAM, CSPG4,* **MMP***1, SPARCL1,* and *MMP10. Down-regulated, IL1B, S100A12, FCGR3B, CCR1, S100A8, CCL3,* **CCL2***, CCL4, CLEC4D,* and *LILRA1.*
Bioinformatics analysis reveals molecular connections between non-alcoholic fatty liver disease [NAFLD] and COVID-19 [2022].	ACE, ADAM17, DPP4, TMPRSS2 and NAFLD-related genes such as **TNF**, **AKT1**, MAPK14, **HIF1A**, SP1, **IL10**.
Organ-specific or personalized treatment for COVID-19: rationale, evidence, and potential candidates [2022].	**CCL2**, **CCL5**, CXCL10, HAO2, BAAT, and SLC27A2.
Differential Co-Expression Network Analysis Reveals Key Hub-High Traffic Genes as Potential Therapeutic Targets for COVID-19 Pandemic [2021].	**IL6**, IL18, **IL10**, **TNF**, SOCS1, **SOCS3**, ICAM1, PTEN, RHOA, GDI2, SUMO1, CASP1, IRAK3, ADRB2, PRF1, GZMB, OASL, **CCL5**, HSP90AA1, HSPD1, IFNG, MAPK1, RAB5A, and TNFRSF1A.
A systems biology approach for investigating significantly expressed genes among COVID-19, hepatocellular carcinoma, and chronic hepatitis B [2022].	ACTB, ATM, CDC42, DHX15, EPRS, GAPDH, **HIF1A**, HNRNPA1, HRAS, HSP90AB1, HSPA8, **IL1B**, **JUN**, POLR2B, PTPRC, RPS27A, SFRS1, SMARCA4, SRC, **TNF**, UBE2I, and VEGFA.
Identification of Key Pathways and Genes in SARS-CoV-2 Infecting Human Intestines by Bioinformatics Analysis [2022]	**AKT1**, TIMP1, NOTCH, CCNA2, **RRM2**, TTK, **BUB1B**, KIF20A, and PLK1.

Note: In bold red, hub genes found in common between different projects.

From these papers, we have collected 142 hub nodes of the liver cells landscape found connected to COVID-19, of which 21.12% comprises a group of 30 genes in common between different projects, while all the others are different. 126 hub genes remain after removing those in common. Barabasi's research consistently showed that biological networks exhibit scale-free properties, with a few genes controlling multiple connections within different functional modules, while most genes have only a few connections [49,50]. It is rather suspicious that the same tissue has a metabolic network operated by such a disproportionate number of hub genes during viral aggression. This suggests heterogeneity of networks. The differences in databases used to extract relationships are a common cause of conflicting results [39,40]. The relationships between the virus and the host occur at the molecular level, mainly through protein interactions. These interactions occur between viral proteins and human proteins and are determined by both human defensive strategies and viral attack strategies. Therefore, it is likely that hub nodes unrelated to the pathology have also been identified. To understand how and why, we applied a biological protocol that involves the identification of the real physical relationships established between the nodes that implement the liver network, with no a priori knowledge of the computational protocols. The fundamental biological events between virus and host drive these interactions, thus necessitating a biological evaluation of each individual interaction (see Methods for details).

Considering the ongoing SARS-CoV-2 pandemic, BioGRID implemented a project called the BioGRID COVID-19 Coronavirus Curation Project. BioGRID is a biomedical interaction repository with experimental data compiled through curation [43]. In these years, BioGRID has accumulated fundamental experimental data supporting the role of SARS-CoV-2 in human infection. This Project collected the comprehensive datasets of all the Known physical interactions between the proteins of the human proteome and those of SARS-CoV-2. In the purification processes of these proteins, researchers overwhelmingly used physical methods such as Affinity Capture-MS and Proximity Label-MS and curators of BioGRID have specifically selected and classified both interactors and physical interactions into various levels of statistical significance. The reason lies because some interactions may be random because the laboratory method does not reproduce the cellular environment. Indeed, the breaking of cells to favor bait-prey interaction also allows for random encounters that do not happen.

Today we have a vast number of over 30 thousand interactions (as of July 2023) from the human proteome when its proteins interact singularly with the 31 viral proteins of SARS-CoV-2. These interactions are unique in being non-redundant and having high confidence interactions at high

throughput, associated with score values of statistical filtering, as determined by using SAINT (Significance Analysis of INTeractome) express version 3.6.0. [51].

We have successfully acquired the entire dataset comprising the entire viral genome (31 proteins) and its interactions with human proteome. With it, we have created a unique database of the human-virus relationships to search for physical/functional interaction between a viral protein and a human protein. Using our proposed conceptual application framework, we can gain a large understanding of the molecular mechanism of a viral infection. A similar approach has already helped researchers recognize targeted viral complexes of five common human viruses [52]. This recognition is based on biological information.

Because of its small genome, a virus must get maximum performance in interfering with the functional processes determined by human cellular proteins aimed at ensuring normal organic homeostasis. The virus learns over time to implement its attack strategy on specific animal targets by evolutionary studying the structure of the target proteins. Many viruses use proteins containing large segments of intrinsic disorder [53] to facilitate "encounter", but every single interaction must have specific and well-defined structural bases to be successful, even if transient. To get this knowledge, the virus employs lengthy periods of co-evolution, parasitizing humans, or similar species [54]. Therefore, if an interaction is present in this peculiar archive, it means that it has a strategic value of attack or defense, for the virus and for humans, respectively. The database also searches for multiple interactions of a human protein with different viral proteins.

Therefore, prioritizing the characterization of the 126 hub genes is an important issue. They should represent the highest-ranking genes, most affected by the virus, and therefore optimal to use as functional seeds. This should significantly facilitate the identification of genes truly associated with the pathology and genes involved in normal metabolic regulation, but also uncertified genes included in networks with no experimental certainty. STRING uses many standardized databases [39] as a source of data and information for calculating network models. It produces a detailed analysis of all the scientific articles underlying each single interaction, and corroborating the models calculated also with biological analyses, such as GO or KEGG, and with structural analyzes using systems such as UniProt. Using STRING, we can manage 6 data channels that parametrize the network calculation differently and influenced by various confidence levels. In this way, we can modulate results with very different parameters of reliability, origin, and statistical significance.

On STRING, we inputted the 126 hub genes as functional seeds to extract their relationships from the entire human proteome. We show this gene list in the Supplements as **Table S1**. These genes, decoded by STRING, should interact to form a protein-protein network model showing also compact sub-graphs. Therefore, we left the six channels open to make the most of all the information from each source, but we set the interaction score to 0.900. As STRING networks usually have a lot of low-scoring interactions, if we want to limit their number per protein, we should use a filter. We used the highest confidence score cut-off to limit the number of interactions to those that have the highest confidence and then are more likely to be true positives. By implementing this strategy, we can narrow down the information only to our input proteins and their network pattern.
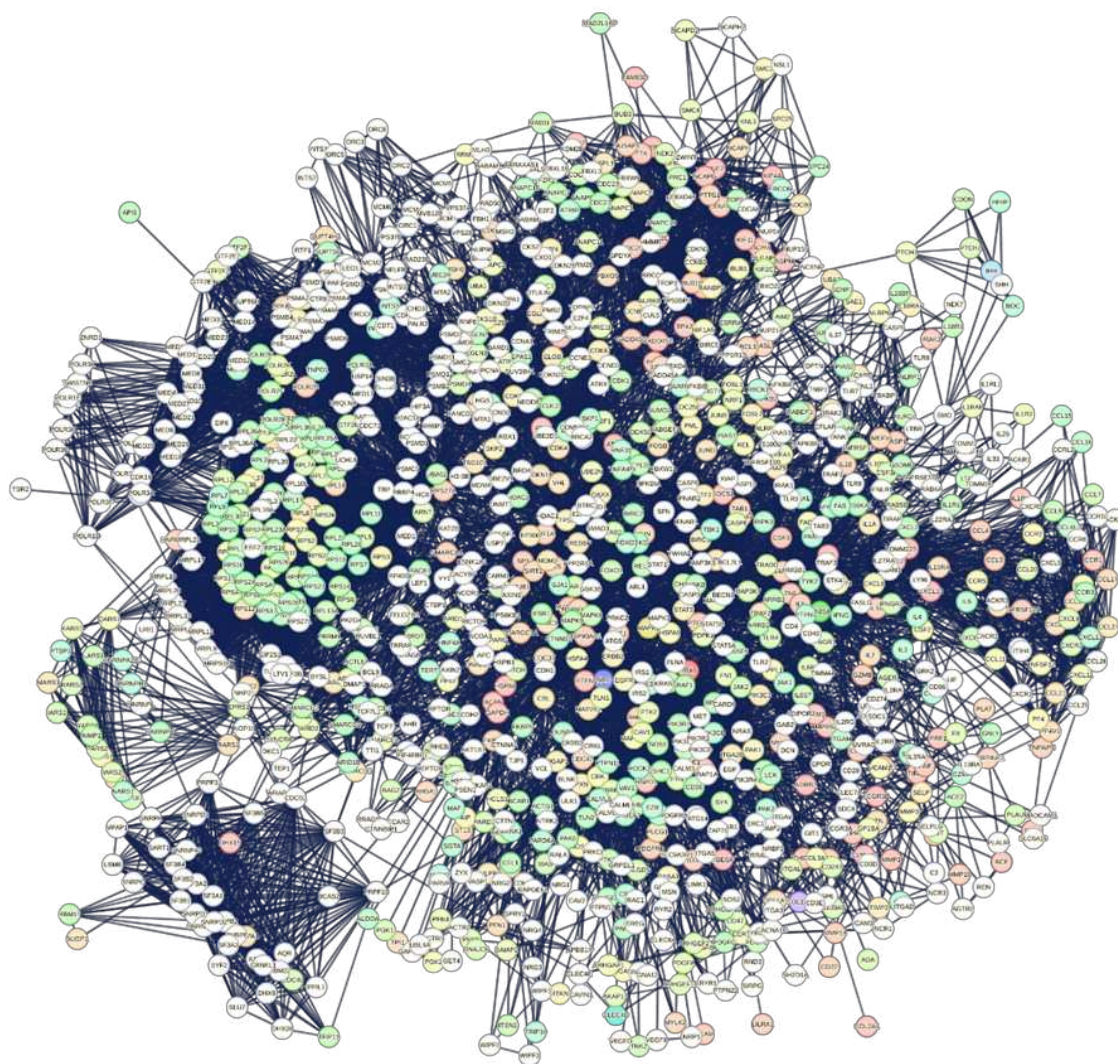
3.2. Comprehensive Liver interactome during COVID-19.

The graph in **Figure S1** of the Supplements shows numerous nodes not connected (31%). A significant number of the remaining elements do not form a compact and connected graph, with only a portion exhibiting connectivity. This is an indicator of poor functional connectivity, but it also says that many of these hubs may not possess the basis of significant experimental certainty. The manipulation of genomic data in the pipeline, from input to the extraction of functional properties of the network, suffers from a lack of accurate data and an indifference for control over know-how. This makes it impossible to carry out any robust analysis, because the disconnected nodes make any topological analysis or functional consideration unreliable [55–58]. To overcome these shortcomings, we can extend the interactions by setting an enrichment of our network with new interaction partners (seeds), always depending on confidence value. This allows us to know whether the input shows evidence of statistical enrichment for any known biological function or pathway. The various external

databases, including Gene Ontology, KEGG pathways, UniProt Keywords, PubMed publications, and others, which annotate the STRING maps, can provide considerable help. The STRING enrichment method retrieves functional enrichment for the set of input proteins. This will show which input protein has enriched terms and the description of each term with all its annotations, providing only answers with FDR =<0.05. About publications, STRING extracts automatically all available scientific texts from PubMed to cover the maximum knowledge about each interaction information, also including full-text articles. **Figure S2** shows the network of **Figure S1** implemented with 500 first-order (direct) nodes and 500 second-order (indirect) nodes. Despite its compactness and size, the resulting graph still shows some unconnected nodes. We removed the 15 unconnected nodes (APOD, BAAT, CCDC112, CSPG4, CYP3A4, DKK3, EPHX4, HAO2, MMP11, NES, PLA2G7, SLC27A2, SPARCL1, STC2, and UGT2B7) using an appropriate tool present in STRING to ensure a fully connected network. Pruning has also the aim of minimizing non-informative enrichment. As a result, we still have 111 residual original hub proteins within the final network, which clearly suggests that we are in the presence of enrichments consistent with the functional seeds used. In **TABLE S2,** we report the list of the 111 remaining hub nodes. It is also important to note that STRING in all the calculated networks has always used data and information extracted from no less than 10,000 scientific articles from PubMed (fully downloadable), which have generated a specific knowledge base for interactions used in the calculation. By employing a sequential cleaning approach, we can get a collection of highly precise information and data, which is ensured by the exceptional dependability of each individual interaction among nodes, unveiling their authentic biological credibility.

The enrichment produced a network that includes all principal human proteins in liver tissues during COVID-19. According to STRING, the net shows 7313 functional associations with biological processes spanning 14 categories. A set of 2344 Biological Processes (GO), 195 KEGG Pathways, and 960 Reactome Pathways characterizes the breadth of functional activities. This network appears very well organized and contains all those functional relationships that also involve the original hub proteins. The compact groupings of certain nodes suggest molecular complexes, even very large ones. We can see these molecular complexes in the peripheral areas of the network. They operate as metabolic nano-machines that carry out specific molecular processes [59,60]. For example, the subgraph at the bottom left is rich in proteins of the Splicing Factor 3B complex that, together with other 17S U2 small nuclear ribonucleoprotein particle (snRNP) components, may play a role in Spliceosome during the selective processing of microRNAs (miRNAs) [61]. This sub-graph also collects many of the proteins involved in transforming the molecules of pre-mRNA (precursor messenger RNA) into mature mRNA. The involvement of this complex is not random because RNA splicing is among the major down-regulated proteomic signatures in COVID-19 patients [62]. Certainly, the virus needs to manipulate the host splicing machinery to its advantage to control the production of its proteome [63]. In fact, going back along the periphery of the network, we encounter compact sets of genes involved in all phases of cellular translational processes and the entire ribosomal complex, just to mention the most important. At least in the liver, these appear to be the most obvious targets of SARS2. The **Excel File 1** reports all the nodes of the interactome in **Figure 1** with their degrees. These nodes also include all the remaining original hubs (111 nodes). In the **Excel file 1,** we can also note a few dozen high-ranking genes, all specific for the various phases of the cytoplasmic translation processes. However, before proceeding with other observations, we have reported in the **Excel file 2** all 26,990 interactions relating to the interactome in **Figure 1**. The file also reports the sources of each single binary interaction and the combined score. The interest in this file released by STRING lies because it shows (in red) the quantitative impact of the component deriving from the experimental data alone on the combined value of the score. Thus, this file is useful as a reference in evaluating each individual interaction for the score of 0.900 (highest confidence) we have always used. As these results show, even for a binary relationship with a score of 0.900, the experimental certification that makes it certain can many times be missing, thus introducing serious and not easily visible anomalies into the graph. We then processed in our SARS2-Human Proteome Interaction Database (SHPID) each single protein of the entire interactome (1111 nodes) to find out

which viral proteins had interacted with the network proteins, as well as with the remaining original hub-proteins. Some of these proteins no longer exhibit the high connectivity characteristics that were crucial when they were designated as HUBs in the original papers. For example, hub nodes like MCAM, LILRA1, GDI2, COL2A1, TNFAIP6 or PTX3 now have low ranks. What happened reveals that their COVID-19-associated high functional rank disappear because they are likely proteins very inflated by high studying frequency because of their relevance in diseases or for their functional importance in the cell or because they are poorly characterized. A quick check using the **Excel file 2** highlighted the widespread lack of valid biochemical and biophysical experimental data for these proteins, meaning that they did not provide adequate evidence for the functional hypotheses in which they had been implicated. Although determining the cellular localization of events through PPI networks is experimentally challenging, in this interactome, we find proteins localize to a precise range of modules, represented by specific molecular complexes.

Figure 1. Comprehensive interactome of liver tissue proteins during COVID-19. STRING calculated the graph through enrichment, using as seeds the set of 111 hub proteins got after pruning. We enriched this network with 500 first-order (direct) nodes and 500 second-order (indirect) nodes. Settings: interaction score of 0.900 (highest confidence); all six channels open. Network parameters: number of nodes, 1111; number of edges, 13,494, while its expected statistical number is 8,838; average node degree, 24.3; avg. local clustering coefficient, 0.623; PPI p-value, <1.0e-16; network diameter, 7; network density, 0.022; network heterogeneity, 1.030; network centralizations, 0.128; connected components, 1. (Topological parameters calculated by Cytoscape).

3.3. Metabolic Stress Related to COVID-19 in the Liver.

The **Excel file 1** shows the protein RPS27A, with a degree of 161, serves as the primary hub. The original hub nodes list (refer to **TABLE 1)** also contained RPS27A. One alias of RPS27A, Ubiquitin-40S Ribosomal Protein S27a, explicitly suggests its function as a remarkably conserved protein responsible for directing cellular proteins toward degradation by the 26S proteasome [64]. Thus, its role in the liver holds significance. It assembles into ribosomes, but also functions independently of them. We also know RPS27A plays a significant role in the progression of various human cancers, including HCC [65]. Its landscape of action during viral infection of the liver is interesting. Investigations of SARS-CoV-2 infection have shown large-scale chromatin structural changes because of metabolic stress [66,67]. In situations of oxidative stress [68], induced by phases of the viral cycle [69,70], both oxidizing agents and the need to signal this stress, as well as variations in sensitivity to oxygen, have highlighted the importance of HIF in signaling [71]. These effects are a common feature of both tumors and COVID-19 [72,73]. In both cases, cells must switch from the TCA cycle to the energetically less efficient glycolysis pathway, and so many glycolytic enzymes are up regulated. One of the transcriptional regulators involved in the response to oxidative stress is HIF1A [74], which remains inactive in normoxic conditions because of its interaction with HIF1AN, an oxygen sensor that hinders interactions with other transcriptional co-activators. SIRT1 serves as an energetic sensor [75], connecting transcriptional regulation to intracellular energetic demands, while TP53BP1 acts as a p53 binding protein, participating in the response to DNA damage.

In tumor progression, the stressful events described affect the p53 protein. The p53 (gene TP53) role is to inhibit the proliferation of cancer cells through cell cycle arrest [76]. Therefore, it normally performs a protective cellular action. The main cellular antagonist of p53 is MDM2, as it triggers the degradation of p53 [77] and effectively supports cancerous growth. MDM2 and p53 establish a feedback loop to preserve balance, complemented by the involvement of RPL11, a ribosomal protein that inhibits MDM2 and enhances p53 stabilization and activation in normal conditions [78]. Therefore, RPL11-MDM2-p53 form an axis regulated precisely by RPS27A [78]. When activated by cellular stress phenomena, RPS27A hinders the interaction between RPL11 and MDM2, promoting the degradation activity of p53 through the catalytic activity of a free MDM2, thus starting the oncogenic process. Hence, this system of proteins works as a sensor and regulator of cellular stress, acting on p53 and RPS27A to regulate their specific activity.

Figure 2 demonstrates the influence of DNA damage and oxidative stress on these same metabolic players during COVID-19. By highlighting the proteins involved in these processes through a tool that colors the nodes specifically involved (refer to Materials and Methods for further information) we can identify them within the liver protein interactome, also visualizing their role and functional relationships. **TABLE 2** shows the activated biological processes, their statistical value, and the colors of the nodes in the network.
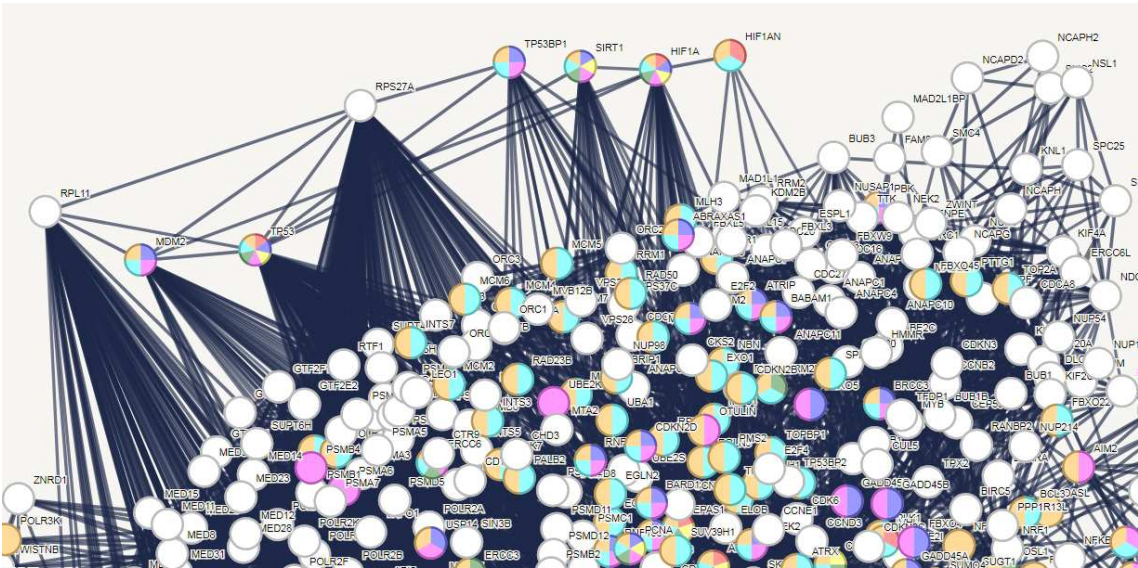
Figure 2. Role of TP53 (p53) and RPS27A in liver infection by SARS-CoV-2. The network is that of **Figure 1** and the nodes at the top left have been carefully extrapolated to highlight both the mutual relationships and the abundance of functional connections with the central core of the network. The degree for each single node is RPL11, 104; MDM2, 45; TP53, 133; RPS27A, 161; TP53BP1, 23; SIRT1, 26; HIF1A, 35; HIF1AN, 5. The colours of the individual nodes show the type of metabolic stress (DNA damage and/or hypoxia) induced by COVID-19 in the liver. The biological stress processes (GO) activated are those shown in **TABLE 2**.

TABLE 2 – BIOLOGICAL PROCESSES RELATED TO COVID METABOLIC STRESS IN THE LIVER			
GO-term Biological Process	**Description**	**p-value**	**Node color**
GO:0043620	Regulation of DNA-templated transcription in response to stress	1.90e-03	🔴
GO:0080135	Regulation to cellular response to stress	9.77e-37	🟣
GO:1900407	Regulation of cellular response to oxidative stress	8.69e-05	🟠
GO:0034599	Cellular response to oxidative stress	1.93e-11	🟡
GO: 0080134	Regulation of response to stress	7.11e-67	🩷
GO:0006979	Response to oxidative stress	2.99e-12	🟢
GO: 0033554	Cellular response to stress	2.98e-45	🔵
GO:0006950	Response to stress	5.86e-86	🟢

The analysis of **Figure 2** reveals that RPL11 and RPS27A are not implicated in the pathways through which cellular stress is detected and transmitted to TP53 and MDM2. These two proteins are not colored; thus, they do not display any stimulation from their interconnected nodes. The non-involvement of RPS27A also suggests that RPL11 continues its activity of blocking the biological function of MDM2 towards TP53. This analysis hypothesizes an activity of TP53 in protecting liver cells by interfering in viral action. Clearly, only data got from laboratory experiments can offer certainties, even though clinical observations of mild liver damage appear to corroborate this hypothesis. However, the **Excel file 2** shows that the experimental component of all the interactions highlighted in **Figure 2** and used to evaluate the hypothesis on the functional activity of TP53 during infection is very high for each protein, so the interactions all rely on solid experimental basis, which strongly supports this conclusion.

3.3. The reverse engineering actions.

The **Excel file 3** reports all the liver proteins that interact with the viral proteins. Only 51 proteins (in red) of the original hubs actually interact with the virus. In our experimental conditions, the human proteins interacting with the 31 viral proteins are only 626 out of 1111 proteins (56%). They originate 2680 interactions with SARS2-host (roughly 20% of the total) of which only 134 can actually be null. These interactions include most of the proteins involved in the translational processes that control protein biosynthesis. In particular, the virus massively takes possession of the ribosomal

system and all the supporting protein complexes to control and promote the biosynthesis of its proteins. This result supports the idea that viruses mostly target high ranked proteins and proteins crucial in certain biological processes [79]. Several authors have already noted this remarkable ability of individual SARS-CoV-2 proteins to interact with many human proteins, drawing therapeutic and pathobiological observations [80–82].

There is a notable difference in action between DNA and RNA viruses. Scientists classify viruses according to their DNA or RNA genome. DNA viruses replicate using DNA-dependent DNA polymerase. RNA viruses exhibit greater heterogeneity, especially with ssRNA (+) viruses like coronaviruses. The genetic material of these viruses is very similar to a mRNA. Compared to the genomes of DNA viruses, RNA viruses have smaller genomes that encode fewer proteins and can undergo rapid and direct translation within the host cell. The proteins of RNA viruses have developed a strategy by interacting with host proteins through specific protein binding motifs. In fact, RNA viruses attacking with few proteins need them to have as multifunctional a capacity as possible. Therefore, we expect RNA virus proteins to possess the capacity to interact with multiple molecular partners. This ability to multitask implies quite specific evolutionary structural adjustments. Indeed, RNA viruses encode proteins characterized by many binding interfaces, but physically with smaller binding surfaces, to hit a greater number of cellular targets [83,84]. Another structural feature to achieve efficient multitasking is to have various segments of intrinsically-disordered-structure along the protein sequences that are very suitable for expressing multiple, even uncorrelated, activities [85,86]. We could say that the proteins of RNA viruses have had a specialized evolution to develop very peculiar biophysical characteristics. It is widely acknowledged that viral nonstructural proteins engage in interactions with host cell proteins, resulting in the formation of replication complexes [87].

Asserting that viral proteins attack human proteins needs quantitative validation and specific information regarding the proteins involved. This question has a particular meaning. In all protein databases, as we have already pointed out, the spatio-temporal characteristics of the archived proteins are missing. The presence of multiple participants hinders the reconstruction of events. While the interaction between many molecules is a recognized concept, the precise mechanisms, meeting sites, timing, and frequency remain elusive. We have limited knowledge in providing mechanistic information about the targeted complex.

3.4. Individual Human proteins interacting with many viral proteins and their distribution graph.

In the Excel 3 file, we can see that some human liver proteins interact with many viral proteins. It is a known fact that multiple viral proteins have the ability to target specific human proteins (90). These interactions described in Excel file 3 could be a resource for researchers aiming to identify important specific host-virus interactions in the dynamics of disease transmission [89]. In particular, to describe the viral diversity associated with different hosts and different tissues, as well as detect shared associations useful for identifying who, where and how they are shared [88,89]. However, some authors report that, in viral infections, the most common ratio of protein-protein interactions between virus and host is 1:1 [90]. Viral proteins, as well as human proteins, are tightly integrated and interact in a specific functional context. This explains much of the binding specificity between proteins. However, even in the best-case scenario, only a handful of viral proteins could interact with a single human protein. This limitation arises from the physical impossibility of locating suitable binding surfaces on a single molecule and the potential electrostatic repulsions and structural constraints caused by proximity on a crowded structure. In the absence of temporal data on the frequency and specificity of these attacks, we can reasonably think that this massive attack is likely directed towards the entire ribosome and its ancillary complexes, of which the targeted protein is a component, given that the most targeted proteins are the ribosomal ones. But this hypothesis also has another side. It shows the total lack, even in the best databases, of the spatio-temporal characteristics relating to individual human proteins. Given the unlikelihood of crowding on a single protein, the attack is more likely to be sequential, i.e., at different times. A comprehensive understanding of

human biology, and that of other living beings, requires acknowledging the dynamic nature of metabolism.

TABLE 3 shows the human proteins most attacked by viral proteins in the range 12 - 20. Its main purpose is to showcase the different levels of affected human proteins, both high **and l**ow. The degree of each protein (see Excel file 1) is in the bracket.   The high degree is justified because the majority are proteins organized into complexes.

Table 3.

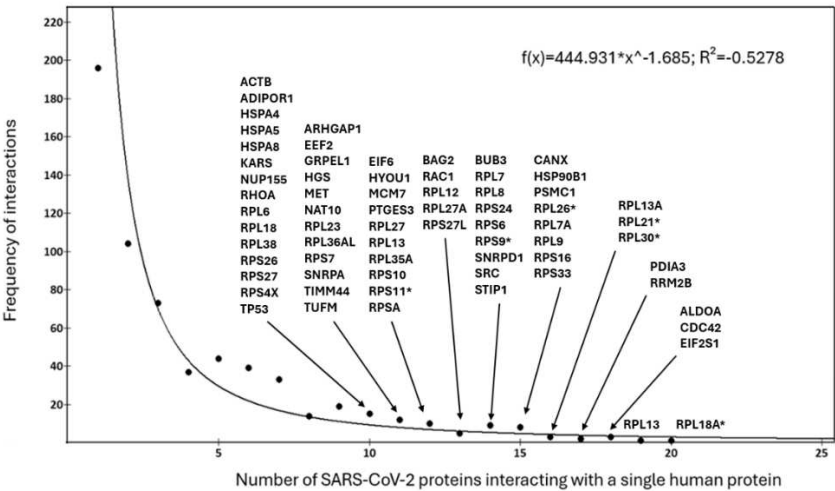| Human Protein	Number of Interacting Viral Proteins**
RPL18A*(84)	20
RPL13(84)	19
ALDOA(4), CDC42(52), EIF2S1(45)	18
RRM2B(3)	17
RPL13A(98), RPL21*(87), RPL30*(85)	16
PSMC1(30), RPL26*(96), RPL7A(85), RPL(9)	15
BUB3(19), RPL7(95), RPL8(95), RPS24(90), RPS6(93), RPS9*(102), SNRPD1(38), SRC(97), STIP1(12)	14
BAG2(7), RAC1(11), RPL12(93), RPL27A(85), RPS27L(82).	13
EIF6(46), MCM7(20), HYOU1, PTGES3(23), RPL27(84), RPL13(84), RPL35A(84), RPS10(87), RPS11*(108), RPSA(99).	12

Note: * Proteins marked with an asterisk also interact with ORF1ab. ** For more extensive details about interactions, see the **EXCEL file 3**.

That some human proteins interact with many viral proteins presupposes many shared structural motifs. But this also suggests that viral motifs in their evolution must gain host-like mechanisms to be successful in invasion. Consequentially, this supports the observations that conformational flexibility, spatial diversity, abundance, and slow evolution are the characteristic features of the human proteins targeted by viral proteins [91]. Viral proteins mimic host binding surfaces of domains to interact with human proteins, which occur through domain-motif interactions. In the **Excel file 3,** we can also observe the interacting viral proteins are not only NSPs (non-structural proteins), but we have also a significant presence of accessory proteins. However, viral proteins intervene in large numbers, targeting mostly the proteins of the ribosomal system. This allows the virus to take control of protein biosynthesis and redirecting it towards the synthesis of the viral genome and its own proteins. That many viral proteins attack one host protein also means that many of them have mimicked the same human motif. In addition, we must consider an average of around 47% of disordered segments in coronavirus proteins [92,93]. This favours attacks on specific cellular targets of the host. An interesting discovery is that among the viral proteins that interact with ribosomal proteins (RPL18A, RPL21, RPL30, RPL26, RPS9, and RPS11) there is also the long viral polypeptide ORF1ab. Since ORF1ab is certainly not a target to be blocked but is the viral polypeptide that must be translated, the asterisked proteins mentioned above could represent points of structural contact of the viral protein ORF1ab with the human ribosome. In fact, some of them (RPL18A, RPL21, RPL30, and RPL26) are specific components of the large ribosomal subunit, the complex responsible for peptide chain elongation and the synthesis of proteins in the cell, while RPS9 and RPS11, are components of the small ribosomal subunit as part of ribosomal process, which couples processing steps of RNA folding, and RNA cleavage [94,95]. Most ribosomes end translation at a stop codon present in the first stem of the pseudo-knot. While, the coronavirus protein-synthesis employs regulatory mechanisms, such as ribosomal frameshifting, promoted by a conserved stem-loop of RNA that forms a promoting pseudo-knot structure [96]. Ribosomes stall at the pseudo-knot and undergo a -1 frameshift at the slippery sequence, leading to the translation of ORF1ab fusion polypeptide [97,98]. In coronavirus, this phenomenon allows the virus to encode multiple types of proteins from a single mRNA, compacting the information. In this way, virus translation dominates host translation because of high levels of virus transcripts.

In **Table 3,** we also find the involvement of lower-degree human proteins that are not ribosomal proteins. Some of them are key because involved in crucial metabolic functions of the liver. We report as examples, ALDOA, RRM2B, BAG2, and HGS. ALDOA is the tetramer of hepatic-type aldolase B that specifically binds to the hepatic cytoskeleton, particularly to actin-containing stress fibers. The presence of disordered segments in the C-terminals favours the possibility of scaffolding and suggests that aldolase can regulate cell contraction [99,100]. RRM2B forms a complex with RRM1 where it plays a key catalytic role in repairing damaged DNA together with p53 and provides deoxyribonucleotides in G1/G2-locked cells [101,102]. While BAG2 is a co-chaperone regulator of the HSP70 and HSC70 chaperones. It acts as a nucleotide exchange factor by promoting the release of ADP from HSP70 and HSC70 proteins, triggering the release of the client/substrate protein [103,104]. In the end, HGS, Hepatocyte Growth Factor, is involved in intracellular signal transduction mediated by cytokines and growth factors. It regulates endosomal sorting and plays a critical role in the recycling and degradation of membrane receptors [105–107]. The liver serves as the site of localization for many of these proteins, emphasizing their tissue specificity.

3.5. Distribution of viral proteins interacting with single human proteins.

Instead, **Figure 3** shows that the distribution graph of the entire set of human liver proteins (626 proteins) interacting with viral proteins (see also **Excel file 3**). Each point on the curve reports the set of human proteins that have the same number of interacting viral proteins. The fit shows that the distribution conforms to a power law, albeit with an R^2 value of 0.5278, suggesting an acceptable fit. This value is at the low limits of reliability and may imply the existence of heterogeneities in the distribution, which makes the results difficult to explain. This should not be surprising because the distribution accurately reflects the overall structural and functional behaviour of the entire set of human proteins with different roles from each other and subjected to sequential functional stress by viral proteins in complex and metabolically differentiated cellular environments. Hundreds of interactions are mainly one-to-one (those on the left side of the curve), while others involve multiple interactions (multi-to-one), to up to 20 viral proteins per single human protein (in the tail). The connectivity distribution in **Figure 3** is quite consistent with the power law's prediction of preferential attachment [108]. Thus, our model should show the emergence of a scale-free topology [109] from interaction results. So, if the connectivity distribution follows a power law, then new nodes will have a better chance of connecting to those with already many neighbours because of the preferential attachment rule.

Figure 3. Distribution of viral proteins interacting with single human proteins. The curve is the exponential fit (displayed at the top right). Data calculated from the **Excel file 3**. The figure also shows, for the experimental points involving the most targeted human proteins (from 10 onwards), the list of which they are. The asterisked proteins are those that also interact with ORF1ab.

Comparative and evolutionary genomic analyses support the birth in the cell of complex structures that make up organized and complicated cellular nano-machines [110]. Genomics has also shown that parts associate with each other to form integrated systems with modular and hierarchical structures [111]. This organizational process should also be intrinsic in the modelling of liver metabolic reactions that arise from protein-protein interactions. In accordance, complex networks exhibit higher-order organization in connectivity, showing links that can be modulated and modelled using subgraphs of the network [112]. Some authors have also shown that networks contain within themselves information about the organization of these compact modules (subgraphs) such as emergence of the protein complexes [113,114]. The peculiarity of these models is the emergence of an important intrinsic structural characteristic of biological networks, namely hierarchical modularity, i.e., a higher level of organization, the growing mechanisms of which, unfortunately, remain unknown. Researchers have never quantitatively tested these qualitative and observational relationships in real biological interaction networks. Our network model, related to liver tissue, shows human protein complexes strongly involved in viral infection. We believe that the preferences of viral proteins toward the interior of these complexes should reflect the mechanisms used by viruses to manipulate host protein complexes.

Based on our collective data, it is evident that the evaluation of virus action should be conducted within the framework of viral preferential attack strategies on intricate protein organizations. However, how viruses manipulate sub-graphs of local host networks, such as human protein complexes, have never been addressed from a topological-computational perspective, preferring to focus on the preferential targeting of viral proteins with hub or bottleneck nodes, despite that no formal definition exists to separate hub proteins from non-hub proteins [12,115].

A systematic analysis of the protein complexes, identified as direct protein-protein targets, has been done to discover new drugs (127) or even through bioinformatic approaches (52), almost never considering a topological point of view. In such type of analysis, both local topological aspects of the network and evolutionary ones should contribute, but, to date, discrimination of the topological and functional properties of complex viral targets during an infection is lacking. Our analysis identified compact sub-networks of human proteins targeted by multiple viral pathogen proteins. But what is perplexing is that during the infection, the targeting process of a complex protein system, such as the ribosome, seems to depend on the connectivity of neighbouring proteins in the network (due to the preferential attachment, which is a topological parameter). Conversely, the interaction of a viral protein ought to be primarily determined by the likelihood of a physical encounter associated with the decrease in free energy because of binding, so exploiting chemical-physical parameters from evolutionary laws.

We can hypothesize, from the analysis presented in **Figure 3**, that multiple types of interaction activities could compete concurrently. If this is the case, upon closer analysis, we should be able to discern more exponential decays that would better characterize the distribution. In **Figure 4 (upper side),** we observe that the degree distribution seems to follow a single power law. But the fit in the log-log scale shows that the single power law distribution is at the lower limit to significantly meet or explain the data characteristics. One-to-one and one-to-many interactions behave differently and make the analytical representation heterogeneous when considered together. The bottom side shows that the distribution, always in the log–log scale, displays two different slopes, unlike what happens when fitting with a single power law. In both fits, the values of R2 are very good, suggesting a combination of two solutions (or two decays) linearly independent. The biphasic distribution opens the hypothesis that there may be at least two dominant classes of coexisting proteins with differentiated functional responses. One class (in black) should contain human proteins essential for metabolic adaptations following viral infection. These proteins can be under-expressed or lost when pathophysiological conditions induce profound metabolic changes. Proteins belonging to the other class (depicted in red) are essential for critical physiological processes of viruses and hosts but are also essential for the virus to gain energy. Thus, these human proteins, highly expressed, exhibit enhanced resistance to pathological processes that induce functional variability. Depending on the characteristics of the local context, it is possible for all proteins to transmigrate between both classes.

In the **lower Figure 4**, there is on the x axis the transition degree, TD. Its value breaks into two parts the distribution and identifies the boundary between nodes with interaction degree less than 12 (in black, made up of proteins that are on average poorly connected), and nodes having degree greater than TD (in red, composed of evolutionarily older proteins that are on average much more connected). In our analysis, each of these sub-networks follow well a single power-law degree distribution, while differing in the value of power-law exponents.
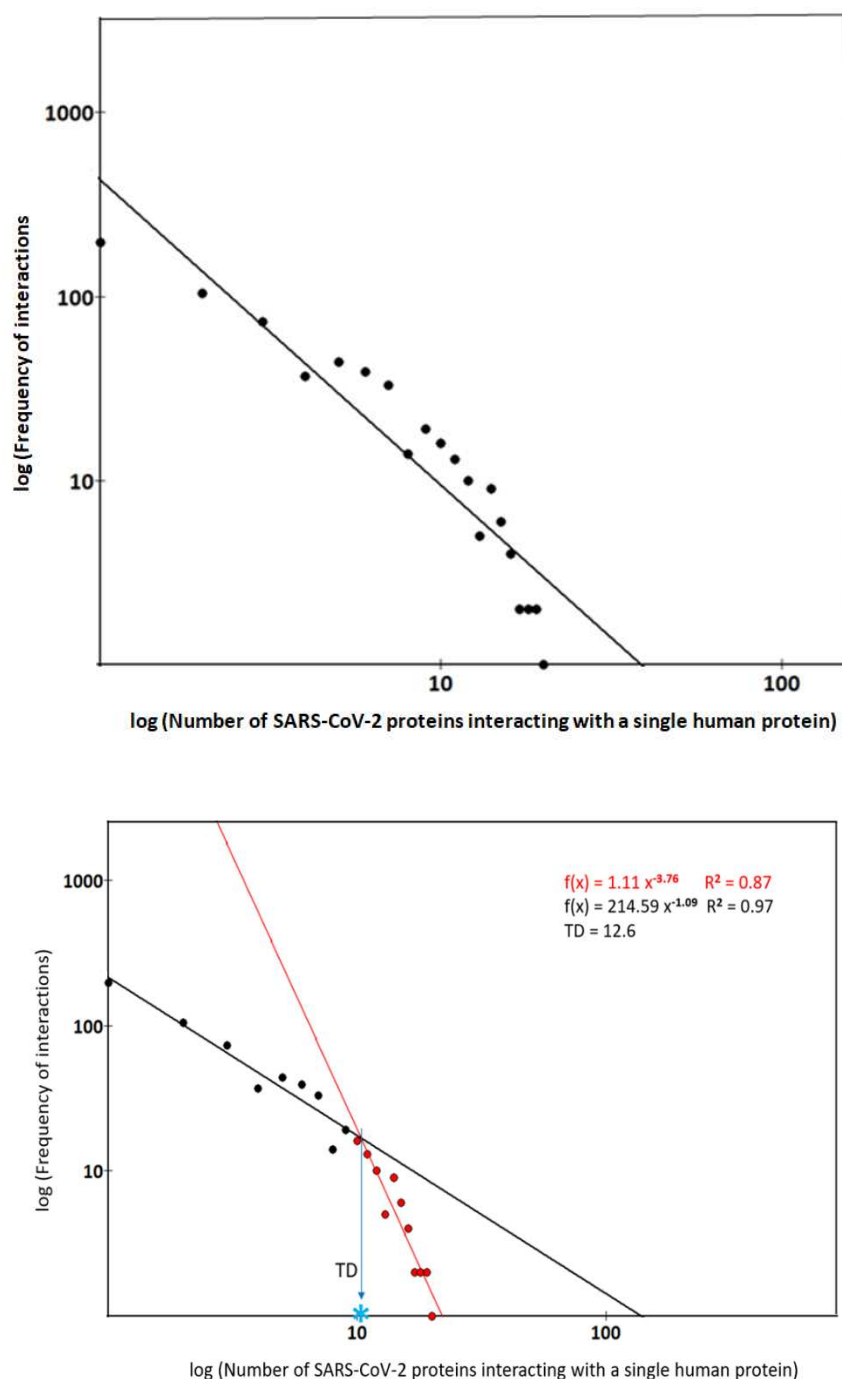
Figure 4. – Linear distributions of interacting viral proteins with a single human protein (log-log scales). **Upper figure** – Distribution graph considered as a single power law. Fitting: $f(x) = 431.26\ x^{-1.66}$ and R^2 is 0.3675. **Lower figure** - Biphasic representation of the power law. The graph displays the fitting equations. TD is the transition degree, the estimated point (marked by blue star) at which the slope of the distribution sharply changes. Its value is around 12.

This biphasic model suggests all proteins can gain new interactions with rate (greater slope) and number of interactions (the rich get richer) always increasing, as happens for older proteins (red ones). Proteins can also lose their interactions, both with and without the loss of their connecting partners. It is a kinetic model which through the different slopes reflects the evolutionary behavior of proteins, considering two classes of proteins, one with a rapid action but also with a fast residence time, the second, with opposite properties of greater resilience. Both classes adequately describe, both in topological and evolutionary terms, the nature of the bi-exponential model. The model, in fact, shows a situation in which the oldest proteins, the most conserved by evolution, increase their interactions because of the establishment of new and specific kinetic conditions. Although our results are built on solid foundations of statistics and experimentation, it is important to interpret them with caution due to all the limitations previously described.

3.6. Comprehensive analysis of liver metabolic activities during COVID-19

To support the structural and functional organizational events previously found for these proteins and the complexes involved, we analyze the data using the many specific databases that STRING maps onto the protein data of calculated networks. **Table 4** reports some analysis of biological processes made by STRING on the interactome data in **Figure 1.** The table shows the most statistically reliable results. Although all data used in this study have a high intrinsic significance, analyzes on extensive sets, where gene expression variability could also play a fundamental role, must be carefully evaluated. Therefore, in their evaluation, the value of the intensity of the expression of the genes that code for the proteins of a process, contained in the Strength parameter (see Methods), was also considered. The results show that the p-value (fdr) is important, but the level of gene expression influences its significance. Then, the intensity of the biological action also depends on the intensity of gene expression.

The gene expression depends on cellular signals, but the biological results depend on the phenotype "interpretation" of that information, which is displayed by the synthesis of proteins (and non-coding RNA). Thus, this parameter allows for the definition of a similarity metric between gene expressions, which we can use to reposition and compare biological processes [128,129].

Table 4.

1 - NORMAL BIOLOGICAL PROCESSES RELATED TO NODES CERTIFIED BY REVERSE ENGINEERING IN THE LIVER INFECTED BY COVID				
GO-term Biological Process	Description	P*	p-value	Strength
GO:0019221	Cytokine-mediated signaling pathway	47.50	8.51e-57	0.82
GO:0002181	Cytoplasmic translation	46.53	2.05e-44	1.05
GO:0071345	Cellular response to cytokine stimulus	42.97	1.59e-63	0.68
GO:0033044	Regulation of chromosome separation	37.73	9.62e-36	1.02
GO:0010965	Regulation of mitotic sister chromatid separation	36.30	8.03e-34	1.04
GO:0033045	Regulation of sister chromatid segregation	36.20	6.46e-34	1.02
GO:0051983	Regulation of chromosome segregation	34.60	4.60e-35	0.97
GO:0030071	Regulation of mitotic metaphase/anaphase transition	33.87	3.68e-32	1.04
GO:0033044	Regulation of chromosome organization	32.37	3.03e-39	0.82
GO:0007346	Regulation of mitotic cell cycle	32.25	1.18e-46	0.70
GO:1901987	Regulation of cell cycle phase transition	30.16	2.98e-42	0.71
GO:0006412	Translation	29.30	4.59e-40	0.72
GO:1901990	Regulation of mitotic cell cycle phase transition	27.66	2.42e-37	0.74
GO:1990869	Cellular response to chemokine	23.92	8.36e-24	0.96
GO:0034243	Regulation of transcript. elongat. from RNA polym. II	17.94	5.25e-19	0.91
GO:0007088	Regulation of mitotic nuclear division	17.50	3.89e-20	0.85
2 - NEGATIVE REGULATION OF BIOLOGICAL PROCESSES RELATED TO NODES CERTIFIED BY REVERSE ENGINEERING IN THE LIVER INFECTED BY COVID				

GO-term Biological Process	Description	P	p-value	Strength
GO:0043069	Negative regulation of programmed cell death	18.94	2.65e-36	0.52
GO:0043066	Negative regulation of apoptotic process	18.31	7.95e-35	0.51
GO:1901988	Negative regulation of cell cycle phase transition	15.97	3.11e-22	0.71
GO:0045786	Negative regulation of cell cycle	15.25	1.63e-24	0.63
GO:0010948	Negative regulation of cell cycle process	14.98	1.08e-22	0.68
GO:0009892	Negative regulation of metabolic process	14.36	3.19e-43	0.33
GO:0010605	Neg. regulation of macromolecule metabolic process	14.22	6.61e-41	0.34
GO:1901991	Neg. regulation of mitotic cell cycle phase transition	13.83	8.82e-18	0.73
GO:0045930	Negative regulation of mitotic cell cycle	13.33	2.12e-19	0.69
GO:0031324	Negative regulation of cellular metabolic process	12.03	2.37e-34	0.35
GO:0060548	Negative regulation of cell death	11.95	1.43e-34	0.35
GO:2000816	Neg. regulation of mitotic sister chromatid separation	11.88	7.56e-11	1.0
GO:0045841	Neg. regulation mitotic metaphase/anaphase transition	10.46	2.29e-10	1.01
GO:2001237	Neg. regulation of extrinsic apoptotic signaling pathway	9.67	5.60e-12	0.76
GO:0051348	Negative regulation of transferase activity	8.90	1.17e-15	0.59

3 - DYSREGULATED BIOLOGICAL PROCESSES RELATED TO NODES CERTIFIED BY REVERSE ENGINEERING IN THE LIVER INFECTED BY COVID

3A -Local network clustering (STRING)**	Description	P	p-value	Strength
CL.152	Viral mRNA Translation	89.03	7.21e-46	1.19
CL:159	Viral mRNA Translation	55.38	1.06e-45	1.23
CL:162	Cytoplasmic ribosomal proteins	54.16	1.41e-43	1.23
CL.143	Viral mRNA Transl. and Sec61 translocon complex	53.10	6.93e-47	1.11

3B - REACTOME PATHWAYS	Description	P	p-value	Strength
HSA-192823	Viral mRNA Translation	64.09	2.56e-53	1.2
HSA-72764	Eukaryotic Translational Termination	61.79	2.32e-52	1.18
HSA-72689	Formation of a pool of free 40S subunits	58.97	1.91e-51	1.15
HSA-72737	CAP-dependent Translation Initiation	53.73	1.98e-49	1.09
HSA-1799339	SRP-dependent cotranslational prot. targeting to membr.	53.17	2.20e-48	1.1
HSA-9679506	SARS-CoV infections	38.58	5.77e-50	0.76
HSA-9754678	SARS-CoV-2 modulation of host translational machinery	26.18	2.39e-23	1.12
HSA-9692914	SARS-CoV-1 host interactions	32.98	1.06e-32	1.03
HSA-9705683	SARS-CoV-2 host interactions	31.14	1.61e-36	0.86
HSA-9678108	SARS-CoV-1 infection	30.73	1.12e-33	0.93
HSA-9735869	SARS-CoV-1 modulates host translational machinery	28.19	1.28e-23	1.22
HAS-9754678	SARS-CoV-2 modulation of host translational machinery	26.18	2.39e-23	1.12
HSA-9694516	SARS-CoV-2 infections	25.52	1.07e-34	0.75
HSA-9705671	SARS-CoV-2 activates/modulates innate/adaptative immune responses	11.06	5.57e-14	0.75
HSA-597592	Post-translational protein modification	7.95	1.28e-22	0.36
HSA-9772572	Early SARS-CoV-2 Infection Events	3.68	1.3e-05	0.72

4 - PROTEIN DOMAIN CHARACTERISTICS IN THE LIVER INFECTED BY COVID

4A – Prot. Domains (InterPro)	Description	Count in network	P	p-value	Strength
IPR036048	Chemokine interleukin-8-like superfamily	29 of 44	15.03	1.11e-14	1.07
IPR039809	Chemokine beta/gamma/delta	15 of 26	8.03	8.90e-07	1.01
IPR033899	CXC Chemokine domain	12 of 14	7.30	1.54e-06	1.18
IPR011332	Zinc-binding ribosomal protein	9 of 10	7.01	6.92e-05	1.2
IPR011029	Death-like domain superfamily	29 of 97	6.01	2.23e-08	0.72
IPR008271	Serine/threonine-protein kinase, active site	52 of 310	4.21	9.00e-08	0.47

IPR001875	Death effector domain	5 of 7	3.84	3.10e-03	1.1
IPR0000488	Death domain	11 of 35	2.81	6.3e-03	0.74
4B - Prot. Domains (SMART)	**Description**	**Count in network**	**P**	**p-value**	**Strength**
SM00199	**Intercrine alpha family (small cyt/chem CXC)**	28 of 42	16.80	5.07e-15	1.07
SM00252	**Src homology 2 domains**	22 of 104	3.24	2.5e-05	0.6
SM00219	**Tyrosine kinase, catalytic domain**	20 of 88	2.64	2.5e-04	0.6
4C - Annotated Keywords (UniProt)	**Description**	**Count in network**	**P**	**p-value**	**Strength**
KW-0689	**Ribosomal protein**	90 of 175	44.83	**5.05e-46**	0.96
KW-0687	**Ribonucleoprotein**	112 / 278	42.17	**4.14e-49**	0.85
KW-0945	**Host-virus interaction**	148 / 540	33.03	**3.81e-48**	0.68
KW-0747	**Spliceosome**	50 of 138	16.77	**5.14e-20**	0.81
KW-0395	**Inflammatory response**	56 of 163	16.56	**1.73e-21**	0.78
KW-0132	**Cell division**	88 of 384	14.25	**2.31e-23**	0.61
KW-0498	**Mitosis**	69 of 75	13.43	**4.53e-20**	0.65
KW-0131	**Cell cycle**	137 / 651	13.13	**1.09e-23**	0.57
KW-0647	**Proteasome**	25 of 52	11.57	**2.74e-12**	0.93

The table is split into four sections that show the primary aspects of the metabolic context encountered by the liver during COVID-19. The data are shown in decreasing order determined by the P value. As we note, some p-values, despite being remarkably low, are repositioned due to variability in the intensity of gene expression. In the first part of the table (**Section 1**) we can see that cellular activity is mainly involved in promoting cytokine signaling processes, cellular translation, and the cell cycle. In the second part (**Section 2**), we have the negative regulations resulting from the viral attack. Surprisingly, one of the main viral activities is to alter the programmed processes of cell death, followed by strong interference to alter the processes of the cell cycle in its various phases. These data suggest a viral activity that aims to implement a systemic spread of intact but infected cells, very similar in result to the spread of cancerous metastases. If we observe the interaction data in **Excel file 3**, we can see that the virus attacks proteins of the cellular matrix and cytoskeleton, such as ACTB, ACTR3, FN1, CDC42, COL2A1, COL18A1, ITGA3, ITGA5, ITGAV, FLNA, ACTL6A, ACTR2/3, and others, similar to what the cancer cell does to spread metastasis. Other researchers have noticed similar strategies [130], such as extending particular stages of the cell cycle and managing programmed cell death. **Section 3A** shows some of clusters calculated by STRING which show the involvement of the virus in mRNA translation and in ribosomal cytoplasmic proteins. Local STRING network clusters are pre-computed protein clusters derived by hierarchically clustering the full STRING network.

The **Supplements** (under Clustering) provide a comprehensive overview of all four clusters of the **Section 3A,** featuring their topological parameters and a GO analysis for each, to facilitate the identification of the metabolic framework of action. Extremely low FDR values characterize all these contexts, demonstrating that the cytoplasmic translational system, including ribosomes, is the most statistically significant virus target.

The **Section 3B** (Reactome) shows the most reliable metabolic pathways that involve extensive virus-host interactions and identifies sets of proteins that also perform the same action as SARS-CoV-1. **Section 4** highlights the specific human protein domains targeted by viruses. One interesting aspect is that they have quantized the presence and incidence (count in the net) of these proteins. Many of these domains (**Sections 4A and 4B**) are involved in the molecular mechanisms of chemokine/cytokine signaling, and in the reprogramming processes of programmed cell death. The last section, 4C, shows in which downregulated biological processes we find these domains and in what abundance, including spliceosome-mediated RNA processing. The set of this information is in excellent agreement with that discussed earlier and also opens up to other observations. Although

our results are built on solid foundations of statistics and experimentation, it is important to interpret them with caution due to all the limitations previously described. In this study, we did not discuss one-to-one interactions of the proteins of this viral pathogen with other human proteins. The most surprising of these observations (see the Excel file 3), is the large number of one-to-one interactions, that, for instance, characterize the S1 viral protein (Spike), which interacts with many individual human proteins involved in different metabolic processes (manuscript in preparation).

4. Discussion

COVID-19 involves many cellular biochemical adaptations affecting specific biochemical and physiological pathways that generate profound systemic alterations which are reflected in specific organ adaptations. This justifies a specific study of the alterations generated in the liver by SARS-CoV-2. The study shows the interactions between viral and human proteins involved in molecular and/or biological processes and their consequences because of the infection. To the best of our knowledge, we have presented here the most comprehensive and in-depth analysis of SARS-CoV-2-Human PPIs within the liver infection by COVID-19.

Our analysis revealed that viral targets are enriched in human protein complexes, such as ribosome or proteasome, and results confirm that viral infection affects large protein complexes involved in the human translational system. During the attack, we observed a significant presence of scaffolding and housekeeping proteins among the viral targets. In this way, the virus takes possession of and controls the entire apparatus that manages mRNA translation, blocking similar activities of the host. The strategy is to encourage viral replication. Therefore, understanding the host molecular mechanisms involved in protein-protein interactions (PPIs) controlled by SARS-CoV-2 is crucial for the design of new antiviral strategies, also because there are human proteins that could be better targets than viral ones. However, the results show the interactions are crucial factors for regulating cellular metabolism and survival during stressing times, which have relevance in viral infections for disease progression.

Many pathological features of SARS-CoV-2 in the liver have remained unclear because the underlying molecular mechanisms are unknown (1). Although many host proteins can interact with viral proteins, only some of them are essential for a full infection in a virus-specific manner. The results also show that the biological control exerted by the various human HUBs, as reported in the literature, was not always confirmed, nor was it showed which of them physically interacted with viral proteins. The results presented in our reverse engineering approach are all experimentally based because the proteins involved and their specific interactions come from BioGRID. Through a comprehensive collection of all BioGRID one-to-one interaction data, we could filter these proteins, revealing the functional characteristics of those involved in virus-host interactions. Although many host proteins can interact with multiple viral proteins, only part of them was crucial for infection in a virus-specific manner, after filtering out the less significant ones to reduce noise.

The limit of this approach does not lie in the methods used, but in the acquisition and representation of tissue information on a spatial and temporal scale, which remains a limit to be overcome technologically. This is the real challenge. Considering the intricacy in representing the spatiotemporal organization of cells and tissues as metabolic scenarios, our aim has been to choose specific biological processes applicable in real-world scenarios. We extracted from the literature an extensive set of heterogeneous hub data of the liver of infected patients by comparing them with the biological data set of our database and pruning those of low significance. We have showed the accuracy and biological robustness of our conclusions. Next, we evaluated these liver datasets and showed they could detect metabolic patterns of hepatic tissues within COVID-19. Our data showed that inverse engineering can map and reconstruct the metabolic distribution of various biomolecules, providing valuable multimodal insights into coronavirus disease.

From the distribution analysis of the human proteins, used as targets by the viral proteins, we have highlighted that the best fit of the data is the one that provides a biphasic power law. This allowed us to highlight at least two classes of proteins related to two different distributions that consider two operational kinetics of the two classes. Evolutionarily consolidated proteins are mainly

connected to one class, which is rather resilient from a functional point of view, increasing their connectivity more quickly. The other class includes proteins that are already weakly connected, essentially more focused on pathological aspects and which respond with a slower growth in connectivity. Thus, the forces driving this protein behavior are both evolutionary and topological, albeit to varying degrees.

A set of over 33 thousand experimental human-virus interactions curated by BioGRID provided the biological basis for motivating each individual interaction. Added to this is that for every single interaction to model the STRING network, we used a score of 0.900. In evaluating key interactions, we have considered the quantitative incidence of the experimental contribution on the value of the combined score using the parametric data reported in the **Excel file 2**. It is worth considering that only a solid experimental basis can make a protein-protein interaction certain and reliable in the real metabolic world. Recent results show that biases of the experimental procedures used to infer networks can strongly affect the resulting topology [116]. We can also expect that study bias can affect the sensitivity of experiments, given that over-studied proteins are tested more frequently than others [117].

Today, a network can capture functional modules and cellular connectivity processes because proteomic data contains a relational and informational component connected to protein-protein interactions. But the biological events that distinguish a cell, whether normal or infected, represent how the genetic code is executed that triggers one of the many metabolic processes of which a hub node is part of or can manage. Therefore, it is practically very difficult, if not impossible, to distinguish when, how, and with whom a hub node is involved in an altered or normal process. As mentioned previously, the actual activity of a node does not derive from understanding the human metabolic activities in which it seems involved, but from a knowledge of the specific spatio-temporal events that involve it. Simply because a key node, both a hub, or bottleneck, is a crossroads through which many and different pieces of information can pass, but we do not know which ones and in what order. This constraint currently limits human knowledge, but we will overcome it to enable drawing real conclusions.

This study analyzes in depth some protein-protein interactions between virus and host involving molecular complexes in the cellular system represented by liver tissue during COVID-19. The results allow us to provide an account, albeit approximate, of the mapping of these interactions. SARS-CoV-2 has identified multi-protein complexes with which high biological functions are associated as optimal targets for attack. An advantage for this virus is that, being a ssRNA (+) virus, it has a very rapid cytoplasmic production of viral proteins. The affected multi-protein complexes are RNA splicing, transcription, and translation machineries, but also cell signaling proteins, which function as part of complexes on the order of mega-Daltons and made of dozens of proteins [116]. With ribosomes and spliceosome, these complexes reach an even greater molecular weight, because, on average, they comprise 100-300 different proteins, including structural and regulatory RNAs [117]. We should also consider that these complexes, which function as scaffolds for viral proteins, are also subject to regulation of their function through the mediation of post-translational modifications. As already noted, we have little knowledge of dynamics about the information flows that drive events that give rise to molecular phenomena, such as signaling or translation. Similarly, we do not know PTMs of subunits and information about the structure/function relationships to organize the architecture of these complexes. All this makes any proposal of a dynamic hypothesis on viral strategy murky. However, although we still have a static understanding of metabolic actions, knowing the details involving some key human proteins in these complexes could open a new era in antiviral pharmacology.

One last observation deserves to be noted to conclude this discussion. We found a smaller quantity of important ribosomal interactions associated with RPLs and RPSs, as opposed to the information documented in the BioGRID file concerning the ORF1ab protein. This result, together with the fact that of the 1111 human proteins of the interactome, only 626 interact with viral proteins, opens considerations on the systemic activity of the virus in various human organs. These results suggest a different viral strategy in different tissues/organs. Many researchers speak of a process of

evolutionary adaptation of the virus to humans, favoured by its successful propensity to mutate quickly. The mutation rate of the virus genome has been estimated of 1×10^{-3} substitutions per base (30 nucleotides/genome) per year under neutral genetic drift conditions or of $1 \times 10^{-5} - 1 \times 10^{-4}$ substitutions per base in each transmission event [118], but, tracking a systematic gene-by-gene comparison analysis with a reference genome (i.e., the first sequence data of a patient from Wuhan in the National Center for Biotechnology Information (NCBI) annotation NC_045512.2), only six of mutations had over 50% frequency in global SARS-CoV-2 up to 2023 (NSP12, S, NSP4, N, ORF9b, and NSP3) [118].

The viral evolution occurs on time scales comparable to virus transmission events and to dynamics that involve many factors [119]. These factors encompass the fluctuation of infected individuals over time, the varying percentages of immune profiles in populations, human mobility, the effectiveness of transmission between individuals, as well as the interplay between viral strains and lineage extinction [119]. The complexity of all this makes it extremely challenging, if not outright impossible, to establish global evolutionary theories through experimental evidence, although it is still workable to have coherent discussions about individual factors of variability. Consequently, numerous hypotheses have emerged regarding the evolution of SARS-CoV-2, including the notion that the virus gradually becomes less virulent. Without going into the merits of these observations and the many existing hypotheses, we note that the sampling of data we collected covers patients scattered around the world who became infected between 2021 and 2023. The genomic profiling focuses solely on the liver. Thus, our data cover a wide window of the evolution of SARS-CoV-2 in relation to liver tissue and regarding high-ranking proteins (hubs), known to be the preferential target of the virus. Although 22% of them did not meet the necessary experimental requirements to be considered reliable, we discovered that only 51 of these proteins (refer to Excel file 3) ultimately played a role in the infection, although many had reduced connectivity. They, through functional enrichment, showed us how remarkable the viral activity was against specific proteins of the entire hepatic cellular translation system. This strategy never changed over 3 years. Checking BioGRID, the interaction data shows that ORF1ab also interacts with many other proteins of the human translational system, but not in the liver. This suggests a unique and specific viral behaviour, i.e., over time, viral methods, and proteins attacking the liver showed no significant changes in strategy. This allows us to hypothesize that it may be reasonable to think of a different strategy regarding the protein-protein interactions of SARS-CoV-2 in the different human tissues/organs. The complexity arises when attempting to illustrate this hypothesis, as the data used is consistently sourced from deceased patients, rendering it impossible to distinguish between the systemic response of the patient's phenotype and the effects specifically tied to the organ being examined. We could also find this information exactly in those poorly interacting hub nodes that often we discard, which could represent unstable ongoing variations of molecular strategy, but not yet consolidated. So, although this result may already exist in another context, where different design objectives obscure it, in this investigation, we present precise molecular data that supports a different way to approach the distribution of nodes in an interactome, opening new design hypotheses. The scientific community should verify these data.

Appendix A

An appendix is necessary to frame the reason for a reverse engineering approach and explain why everything must be based on reliable data. When we think of a biological network, we think it comprises a one-to-one set of interactions of its nodes. What the nodes exchange is functional information, therefore, a biological network is an information system that manages the metabolic information of the cell and the entire organism. The more precise the metabolic information, the greater the homeostatic capacity of the entire organism. Therefore, when we refer to two nodes that exchange functional relationships, i.e., information, we must be very sure that the interaction exists. We can only get high certainty experimentally, for example, through the methods of biochemistry and biophysics. The mutual information between two variables, i.e., the nodes, measures the amount of information that one variable contains about another. The more certain and higher the certainty, the greater the reduction of the functional uncertainty of one variable following the knowledge of another. Mutual information between two variables is a fundamental concept of information theory as defined by Shannon [36]. To apply this concept to one-to-one biological interactions, we should define it in terms of entropy, as an uncertainty of the information that is transmitted [37].

Reciprocal information is a measure of dependencies between variables, which can analyze interaction networks: if two components have strong interactions, their reciprocal information will be high, increasing the certainty of the event. The intrinsic information of an event, also called self-information, is the amount of intrinsic uncertainty associated with it. The more certain the event, the lower the amount of uncertainty associated with it. From which the more certain the information one possesses, knowing that the event has occurred, the lower the associated uncertainty will be, but the lower the self-information or intrinsic information, i.e. its total entropy, will also be.

In entropic terms:

$$I(X, Y) = H(X) – H(X/Y) \tag{1}$$

Where, I(X, Y) is the uncertainty existing on the relationship between the two nodes X and Y and depends on the level of informational uncertainty relating to each variable. H(X) is the entropy of the information system, or self-information or intrinsic information, which is the amount of uncertainty associated with the interaction between the two nodes. H(X/Y) is the conditional entropy, i.e., the entropy of a variable Y conditioned from the knowledge we have of the other variable, the higher the knowledge, the lower the associated uncertainty.

Relation 1] tells us that the uncertainty between two metabolic nodes, which exchange a physical and/or functional relationship [I(X, Y)], depends on the intrinsic uncertainty [H(X)] associated with the relationship itself. We can reduce this uncertainty by increasing our knowledge of the metabolic behaviour of the interacting nodes [H(X/Y)]. However, it definitely states that by confirming a metabolic event between two nodes through experimentation, we get secure and get certain information, thus eliminating the conditional uncertainty and reducing the self-information, intrinsic information, or entropy of my system.

It is crucial to acknowledge that information fully controls biological networks. The greater the certainty of the information we possess about the physical/functional relationships existing in the network, the more certain and real the metabolic or pathological previsions of our computational model are. Therefore, since the relationships between two metabolic nodes are physical/functional, we gain the greatest certainty of all only through conducting experiments, with the methods of biophysics, with which we measure the type and strength of the interaction, and of biochemistry, with which we measure the levels of function. So, the relationships between computational models and experimental data are one cornerstone of systems biology. Reverse engineering aims to understand which functional processes are real and which are dysregulated through external control of the certainty of the biological event of virus-host interaction. The goal is the ultimate biological determination of existing interactions, not the detailed characterization of these interactions, knowing that difficulties increase because we deal with nonlinear interactions.

The presence of many errors undermines these principles very often because uncertainty is intrinsic in the multiple contexts that provide data and information relating to the biomolecules necessary to calculate biological networks. Although next-generation sequencing studies provide

extensive sequence information, the precise knowledge of virus-host one-to-one protein interactions and potential targets for antiviral therapies remains limited, partial, and incomplete. Typically, metadata for PPIs [38] should include experimental details of tens of thousands of virus-human interactions. Some databases, such as BioGRID, STRING, or INTACT, have used standardized procedures, but many others, generalists, have collected virus-host interactions in different ways and contexts [39] and do not have a standard format.

These platforms are online and useful for checking results. The fundamental reason lies in the distinction between reproducibility (repeating an experiment to get the same result) and replicability (interpreting the same data in different contexts) is crucial. It is important to recognize that interpretations of data may vary depending on context, data quality, or analysis method. Standardization of data and protocols is necessary to get a univocal understanding and interpretation of research results. The vast differences between databases make it extremely challenging to compare their data, particularly when the lack of experimental details obscures the nature of an interaction. What we often observe in the interactomics papers is an abnormal bloom of hub genes/proteins far beyond the needs of any biological network [40]. Therefore, the use of STRING, a platform that for each calculated interaction in a graph creates a specific knowledge base by querying thousands of scientific articles on PubMed, and BioGRID, a platform that archives only curated experimental data of the one-to-one interactions of SARS-CoV-2 proteins with the human proteome, are two indispensable tools to guarantee the best possible certainty of the data under analysis. The liver is a very complex organ with a highly dynamic metabolism, where the sequential regulation of cellular processes plays a crucial role [41]. Therefore, studying its metabolic behavior during COVID-19 requires knowledge of the control systems and areas [42], which is not always found out in liver diseases [40].

References

1. Kariyawasam JC, Jayarajah U, Abeysuriya V, Riza R, Seneviratne SL. Involvement of the Liver in COVID-19: A Systematic Review. Am J Trop Med Hyg. 2022 Feb 24;106(4):1026–41. doi: 10.4269/ajtmh.21-1240. Epub ahead of print. PMID: 35203056; PMCID: PMC8991364.
2. Beigmohammadi MT Jahanbin B Safaei M Amoozadeh L Khoshavi M Mehrtash V Jafarzadeh B Abdollahi A , 2021. Pathological findings of postmortem biopsies from lung, heart, and liver of 7 deceased COVID-19 patients. Int J Surg Pathol 29: 135–145. [PMC free article] [PubMed] [Google Scholar].
3. Ryan, Paul MacDaragh, and Noel M. Caplice. "Is adipose tissue a reservoir for viral spread, immune activation, and cytokine amplification in coronavirus disease 2019." Obesity 28.7 (2020): 1191-1194.
4. Hamming, Inge, et al. "Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis." The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland 203.2 (2004): 631-637.
5. Ding, Yanqing, et al. "Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: implications for pathogenesis and virus transmission pathways." The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland 203.2 (2004): 622-630.
6. Birman, Dinami. "Investigation of the Effects of COVID-19 on Different Organs of the Body." Eurasian Journal of Chemical, Medicinal and Petroleum Research 2.1 (2023): 24-36.
7. Paolini, Annamaria, et al. "Cell death in coronavirus infections: Uncovering its role during COVID-19." Cells 10.7 (2021): 1585.
8. Yuan, Cui, et al. "The role of cell death in SARS-CoV-2 infection." Signal Transduction and Targeted Therapy 8.1 (2023): 357.
9. Jothimani, Dinesh, et al. "COVID-19 and the liver." Journal of hepatology 73.5 (2020): 1231-1240.
10. Guan, Guan-Wen, et al. "Exploring the mechanism of liver enzyme abnormalities in patients with novel coronavirus-infected pneumonia." Chinese journal of hepatology 28.2 (2020): 100-106.
11. Shi, Jihang, et al. "Exploration and verification of COVID-19-related hub genes in liver physiological and pathological regeneration." Frontiers in Bioengineering and Biotechnology 11 (2023): 1135997.
12. Vandereyken, Katy, et al. "Hub protein controversy: taking a closer look at plant stress response hubs." Frontiers in plant science 9 (2018): 694.
13. Huang, Tengda, et al. "Demonstration of the impact of COVID-19 on metabolic associated fatty liver disease by bioinformatics and system biology approach." Medicine 102.35 (2023): e34570.

14. Luo, Huiyan, et al. "Comprehensive DNA methylation profiling of COVID-19 and hepatocellular carcinoma to identify common pathogenesis and potential therapeutic targets." Clinical Epigenetics 15.1 (2023): 100.

15. Shi, Jihang, et al. "Exploration and verification of COVID-19-related hub genes in liver physiological and pathological regeneration." Frontiers in Bioengineering and Biotechnology 11 (2023): 1135997.

16. Jiang, Shi-Tao, et al. "Systems biology approach reveals a common molecular basis for COVID-19 and non-alcoholic fatty liver disease (NAFLD)." European journal of medical research 27.1 (2022): 1-21.

17. Shen, Qinyan, Jiang Wang, and Liangying Zhao. "To investigate the internal association between SARS-CoV-2 infections and cancer through bioinformatics." Mathematical Biosciences and Engineering: MBE 19.11 (2022): 11172-11194.

18. Wang, Luhong, et al. "Target and drug predictions for SARS-CoV-2 infection in hepatocellular carcinoma patients." Plos one 17.5 (2022): e0269249.

19. Abolfazli, Pouria, et al. "Bioinformatics analysis reveals molecular connections between non-alcoholic fatty liver disease (NAFLD) and COVID-19." Journal of Cell Communication and Signaling 16.4 (2022): 609-619.

20. Mousavi, Seyedeh Zahra, Mojdeh Rahmanian, and Ashkan Sami. "Organ-specific or personalized treatment for COVID-19: rationale, evidence, and potential candidates." Functional & Integrative Genomics 22.3 (2022): 429-433.

21. Hasankhani, Aliakbar, et al. "Differential co-expression network analysis reveals key hub-high traffic genes as potential therapeutic targets for COVID-19 pandemic." Frontiers in Immunology 12 (2021): 789317.

22. Sokouti, Babak. "A systems biology approach for investigating significantly expressed genes among COVID-19, hepatocellular carcinoma, and chronic hepatitis B." Egyptian Journal of Medical Human Genetics 23.1 (2022): 146.

23. Chen, Ji-Chun, et al. "Identification of key pathways and genes in SARS-CoV-2 infecting human intestines by bioinformatics analysis." Biochemical Genetics (2021): 1-19.

24. Steuer, Ralf. "Computational approaches to the topology, stability and dynamics of metabolic networks." Phytochemistry 68.16-18 (2007): 2139-2151.

25. Hartman, Erik, et al. "Interpreting biologically informed neural networks for enhanced proteomic biomarker discovery and pathway analysis." Nature Communications 14.1 (2023): 5359.

26. Wu, Shuang, et al. "The metabolomic physics of complex diseases." Proceedings of the National Academy of Sciences 120.42 (2023): e2308496120.

27. Yang Y, Fang Q, Shen HB. Predicting gene regulatory interactions based on spatial gene expression data and deep learning. PLoS Comput Biol. 2019 Sep 17;15(9):e1007324. doi: 10.1371/journal.pcbi.1007324. PMID: 31527870; PMCID: PMC6764701.

28. Chikofsky, Elliot J., and James H. Cross. "Reverse engineering and design recovery: A taxonomy." IEEE software 7.1 (1990): 13-17.

29. Fornito, Alex, Andrew Zalesky, and Michael Breakspear. "Graph analysis of the human connectome: promise, progress, and pitfalls." Neuroimage 80 (2013): 426-444.

30. Green, Sara. "Can biological complexity be reverse engineered?." Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 53 (2015): 73-83.

31. Natale, Joseph L., et al. "Reverse-engineering biological networks from large data sets." arXiv preprint arXiv:1705.06370 (2017).

32. de Camargo, Ricardo Saraiva, Gilberto de Miranda, and Arne Løkketangen. "A new formulation and an exact approach for the many-to-many hub location-routing problem." Applied Mathematical Modelling 37.12-13 (2013): 7465-7480.

33. Yimiao Qu, et al., Non-epigenetic mechanisms enable short memories of the environment for cell cycle commitment. bioRxiv **2020**.08.14.250704; doi: https://doi.org/10.1101/2020.08.14.250704

34. Angela Oliveira Pisco, et al., Conceptual Confusion: The case of Epigenetics. bioRxiv 053009; doi: https://doi.org/10.1101/053009.

35. Squire, Larry R., et al. "Memory consolidation." Cold Spring Harbor perspectives in biology 7.8 (2015): a021766.

36. Shannon C. 1948 A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423. (doi:10.1002/j.1538-7305.1948.tb01338.x).

37. Cover T& Thomas J. 1991 Elements of information theory. New York, NY: Wiley.

38. Orchard S. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. Nat. Methods. 2012;9(4):345–350.

39. Szklarczyk, Damian, and Lars Juhl Jensen. "Protein-protein interaction databases." Protein-Protein Interactions: Methods and Applications (2015): 39-56.

40. Sharma A, and Colonna G. "System-Wide Pollution of Biomedical Data: Consequence of the Search for Hub Genes of Hepatocellular Carcinoma Without Spatiotemporal Consideration." Mol Diagn Ther. 2021 Jan;25(1):9-27. doi: 10.1007/s40291-020-00505-3. Epub 2021 Jan 21.

41. Tyson JJ, Chen KC& Novak B. 2003 Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr. Opin. Cell Biol. 15, 221–231. (doi:10.1016/S0955-0674(03)00017-6).

42. Kremling A& Saez-Rodriguez J. 2007 Systems biology—an engineering perspective. J. Biotechnol. 129, 329–351. (doi:10.1016/j.jbiotec.2007.02.009).

43. Oughtred R, et al., The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. 2021 Jan;30(1):187-200. doi: 10.1002/pro.3978. Epub 2020 Nov 23. PMID: 33070389; PMCID: PMC7737760.

44. Szklarczyk D, et al., The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021 Jan 8;49(D1):D605-D612. doi: 10.1093/nar/gkaa1074. Erratum in: Nucleic Acids Res. 2021 Oct 11;49(18):10800. PMID: 33237311; PMCID: PMC7779004.

45. Szklarczyk D, et al., The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638-D646. doi: 10.1093/nar/gkac1000. PMID: 36370105; PMCID: PMC9825434.

46. Doncheva NT, et al., Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J Proteome Res. 2019 Feb 1;18(2):623-632. doi: 10.1021/acs.jproteome.8b00702. Epub 2018 Dec 5. PMID: 30450911; PMCID: PMC6800166.

47. Chung F, Lu L, Dewey TG, Galas DJ. Duplication models for biological networks. J Comput Biol. 2003;10(5):677-87. doi: 10.1089/106652703322539024. PMID: 14633392.

48. Scardoni G, Tosadori G, Faizan M, Spoto F, Fabbri F, Laudanna C. Biological network analysis with CentiScaPe: centralities and experimental dataset integration. F1000Res. 2014 Jul 1;3:139. doi: 10.12688/f1000research.4477.2. PMID: 26594322; PMCID: PMC4647866.

49. Wuchty, Stefan, Erszébet Ravasz, and Albert-László Barabási. "The architecture of biological networks." Complex systems science in biomedicine (2006): 165-181.

50. Almaas, Eivind, Alexei Vázquez, and Albert-László Barabási. "Scale-free networks in biology." Biological networks 3.1 (2007).

51. Teo, Guoci, et al. "SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software." Journal of proteomics 100 (2014): 37-43.

52. Yang S, et al., Understanding Human-Virus Protein-Protein Interactions Using a Human Protein Complex-Based Analysis Framework. mSystems. 2019 Apr 9;4(2):e00303-18. doi: 10.1128/mSystems.00303-18. PMID: 30984872; PMCID: PMC6456672.

53. Mishra, Pushpendra Mani, et al. "Intrinsically disordered proteins of viruses: Involvement in the mechanism of cell regulation and pathogenesis." Progress in molecular biology and translational science 174 (2020): 1-78.

54. Villarreal, Luis P. "The widespread evolutionary significance of viruses." Origin and Evolution of Viruses (2008): 477-516.

55. Guidotti, R., et al., Network reliability analysis with link and nodal weights and auxiliary nodes, Structural Safety, Volume 65, 2017, Pages 12-26. ISSN 0167-4730, https://doi.org/10.1016/j.strusafe.2016.12.001.

56. De Vico Fallani F, et al., Graph analysis of functional brain networks: practical issues in translational neuroscience. Philos Trans R Soc Lond B Biol Sci. 2014 Oct 5;369(1653):20130521. doi: 10.1098/rstb.2013.0521. PMID: 25180301; PMCID: PMC4150298.

57. V. Li and J. Silvester, "Performance Analysis of Networks with Unreliable Components," in IEEE Transactions on Communications, vol. 32, no. 10, pp. 1105-1110, October 1984, doi: 10.1109/TCOM.1984.1095978.

58. S. Knight, H. X. Nguyen, N. Falkner, R. Bowden and M. Roughan, "The Internet Topology Zoo," in IEEE Journal on Selected Areas in Communications, vol. 29, no. 9, pp. 1765-1775, October 2011, doi: 10.1109/JSAC.2011.111002.

59. Militello, G., and Álvaro M. "Structural and organisational conditions for being a machine." Biology & Philosophy 33.5-6 (2018): 35.

60. Akyildiz, I.F., et al., "Nanonetworks: A new communication paradigm." Computer Networks 52.12 (2008): 2260-2279.

61. Will CL, et al., Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. EMBO J. 2002 Sep 16;21(18):4978-88. doi: 10.1093/emboj/cdf480. PMID: 12234937; PMCID: PMC126279.

62. Wang C, et al., Abnormal global alternative RNA splicing in COVID-19 patients. PLoS Genet. 2022 Apr 14;18(4):e1010137. doi: 10.1371/journal.pgen.1010137. PMID: 35421082; PMCID: PMC9089920.

63. Wang ET, et al., Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456(7221):470–6. doi: 10.1038/nature07509.

64. Luo J, Zhao H, Chen L, Liu M. Multifaceted functions of RPS27a: An unconventional ribosomal protein. J Cell Physiol. 2023 Mar;238(3):485-497. doi: 10.1002/jcp.30941. Epub 2022 Dec 29. PMID: 36580426.

65. Zhang, Y., et al., Polymeric immunoglobulin receptor (PIGR) exerts oncogenic functions via activating ribosome pathway in hepatocellular carcinoma. International Journal of Medical Sciences, 2021, 18(2), 364–371. https://doi.org/10.7150/ijms.49790.

66. Vandelli, A., et al. "Structural analysis of SARS-CoV-2 genome and predictions of the human interactome." Nucleic acids research 48.20 (2020): 11270-11283.

67. Chiariello, A.M., et al. "Multiscale modelling of chromatin 4D organization in SARS-CoV-2 infected cells." bioRxiv (2023).

68. Chernyak, B. V., et al. "COVID-19 and oxidative stress." Biochemistry (Moscow) 85 (2020): 1543-1553.

69. Jana, Sirsendu, et al. "HIF-1α-Dependent Metabolic Reprogramming, Oxidative Stress, and Bioenergetic Dysfunction in SARS-CoV-2-Infected Hamsters." International Journal of Molecular Sciences 24.1 (2022): 558.

70. Serebrovska, Z.O., et al. "Hypoxia, HIF-1α, and COVID-19: from pathogenic factors to potential therapeutic targets." Acta Pharmacologica Sinica 41.12 (2020): 1539-1546.

71. Wing, Peter AC, et al. "Hypoxic and pharmacological activation of HIF inhibits SARS-CoV-2 infection of lung epithelial cells." Cell reports 35.3 (2021).

72. Zhu, Z., et al., "Comparison of COVID-19 and lung cancer via reactive oxygen species signaling." Frontiers in Oncology 11 (2021): 708263.

73. Bhandari, V., et al. "Molecular landmarks of tumor hypoxia across cancer types." Nature genetics 51.2 (2019): 308-318.

74. Cimmino, Fl., et al. "HIF-1 transcription activity: HIF1A driven response in normoxia and in hypoxia." BMC medical genetics 20 (2019): 1-15.

75. Varghese, B., et al. "SIRT1 activation promotes energy homeostasis and reprograms liver cancer metabolism." Journal of Translational Medicine 21.1 (2023): 627.

76. Wang, X., et al., "p53: protection against tumor growth beyond effects on cell cycle and apoptosis." Cancer research 75.23 (2015): 5001-5007.

77. Moll, Ute M., and Oleksi Petrenko. "The MDM2-p53 interaction." Molecular cancer research 1.14 (2003): 1001-1008.

78. Liu, Y., et al., "RP–MDM2–p53 pathway: Linking ribosomal biogenesis and tumor surveillance." Trends in cancer 2.4 (2016): 191-204.

79. Rachita R., et al., Molecular principles of human virus protein–protein interactions Bioinformatics 31 (7), 2014, 11-21, 1025-1033, Oxford University Press.

80. Gordon DE et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583, 459–468 (2020).

81. Gordon DE et al. Comparative host–coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science 370, eabe9403 (2020).

82. Komarova, A.V., et al. "Identification of RNA partners of viral proteins in infected cells." RNA biology 10.6 (2013): 943-956.

83. Li J et al. Virus–host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. Med. 2, 99–112 (2021).

84. Stukalov A et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. Nature 594, 246–252 (2021).

85. Zhou Y, et al., A comprehensive SARS-CoV-2-human protein-protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets. Nat Biotechnol. 2023 Jan;41(1):128-139. doi: 10.1038/s41587-022-01474-0. Epub 2022 Oct 10. PMID: 36217030; PMCID: PMC9851973.

86. Khorsand B, et al., SARS-CoV-2-human protein-protein interaction network. Inform Med Unlocked. 2020;20:100413. doi: 10.1016/j.imu.2020.100413. Epub 2020 Aug 13. PMID: 32838020; PMCID: PMC7425553.

87. Ghosh N, et al., Interactome of human and SARS-CoV-2 proteins to identify human hub proteins associated with comorbidities. Comput Biol Med. 2021 Nov; 138:104889. doi: 10.1016/j.compbiomed.2021.104889. Epub 2021 Oct 6. PMID: 34655901; PMCID: PMC8492901.

88. Srinivasan, S.; et al., Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins. Viruses 2020, 12, 360. https://doi.org/10.3390/v12040360

89. Shuler, Gal, and Tzachi Hagai. "Rapidly evolving viral motifs target biophysically constrained binding pockets of host proteins." bioRxiv (2022): 2022-01.

90. Mendez-Rios J, Uetz P. Global approaches to study protein-protein interactions among viruses and hosts. Future Microbiol. 2010 Feb;5(2):289-301. doi: 10.2217/fmb.10.7. PMID: 20143950; PMCID: PMC2832059.

91. Halehalli, Rachita Ramachandra, and Hampapathalu Adimurthy Nagarajaram. "Molecular principles of human virus protein–protein interactions." Bioinformatics 31.7 (2015): 1025-1033.

92. Goh, G.K.M., Dunker, A.K. and Uversky V.N. "Understanding viral transmission behavior via protein intrinsic disorder prediction: Coronaviruses." Journal of pathogens 2012 (2012).

93. Anjum, Farah, et al. "Identification of intrinsically disorder regions in non-structural proteins of SARS-CoV-2: New insights into drug and vaccine resistance." Molecular and Cellular Biochemistry 477.5 (2022): 1607-1619.

94. Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R. Structures of the human and Drosophila 80S ribosome. Nature. 2013 May 2;497(7447):80-5. doi:10.1038/nature12104. PMID: 23636399.

95. Singh S, Vanden Broeck A, Miller L, Chaker-Margot M, Klinge S. Nucleolar maturation of the human small subunit processome. Science. 2021 Sep 10;373(6560):eabj5338. doi: 10.1126/science.abj5338. Epub 2021 Sep 10. PMID: 34516797; PMCID: PMC8744464.

96. Baranov PV, et al., (February 2005). "Programmed ribosomal frameshifting in decoding the SARS-CoV genome" Virology. 332 (2): 498–510. doi:10.1016/j.virol.2004.11.038

97. Rehfeld, Frederick, et al. "CRISPR screening reveals a dependency on ribosome recycling for efficient SARS-CoV-2 programmed ribosomal frameshifting and viral replication." Cell Reports 42.2 (2023).

98. Khrustalev, Vladislav Victorovich, et al. "Translation-associated mutational U-pressure in the first ORF of SARS-CoV-2 and other coronaviruses." Frontiers in Microbiology 11 (2020): 559165.

99. Kusakabe T, et al., Mode of interactions of human aldolase isozymes with cytoskeletons. Arch Biochem Biophys. 1997 Aug 1;344(1):184-93. doi: 10.1006/abbi.1997.0204. PMID: 9244396.

100. Esposito G, et al., Human aldolase A natural mutants: relationship between flexibility of the C-terminal region and enzyme function. Biochem J. 2004 May 15;380(Pt 1):51-6. doi: 10.1042/BJ20031941. PMID: 14766013; PMCID: PMC1224144.

101. Guittet O, et al., Mammalian p53R2 protein forms an active ribonucleotide reductase in vitro with the R1 protein, which is expressed both in resting cells in response to DNA damage and in proliferating cells. J Biol Chem. 2001 Nov 2;276(44):40647-51. doi: 10.1074/jbc.M106088200.

102. Yamaguchi T, et al., p53R2-dependent pathway for DNA synthesis in a p53-regulated cell cycle checkpoint. Cancer Res. 2001 Nov 15;61(22):8256-62. PMID: 11719458.

103. Rauch JN, Gestwicki JE. Binding of human nucleotide exchange factors to heat shock protein 70 (Hsp70) generates functionally distinct complexes in vitro. J Biol Chem. 2014 Jan 17;289(3):1402-14. doi: 10.1074/jbc.M113.521997. Epub 2013 Dec 5. PMID: 24318877; PMCID: PMC3894324.

104. Takayama S, et al., An evolutionarily conserved family of Hsp70/Hsc70 molecular chaperone regulators. J Biol Chem. 1999 Jan 8;274(2):781-6. doi: 10.1074/jbc.274.2.781. PMID: 9873016.

105. Yu Z, et al., Hepatocyte growth factor-regulated tyrosine kinase substrate is essential for endothelial cell polarity and cerebrovascular stability. Cardiovasc Res. 2021 Jan 21;117(2):533-546. doi: 10.1093/cvr/cvaa016. PMID: 32044971; PMCID: PMC7820882.

106. Wu L, et al., O-GlcNAcylation regulates epidermal growth factor receptor intracellular trafficking and signaling. Proc Natl Acad Sci U S A. 2022 Mar 8;119(10):e2107453119. doi: 10.1073/pnas.2107453119. Epub 2022 Mar 3. PMID: 35239437; PMCID: PMC8915906.

107. Han J, et al., Involvement of CASP9 (caspase 9) in IGF2R/CI-MPR endosomal transport. Autophagy. 2021 Jun;17(6):1393-1409. doi: 10.1080/15548627.2020.1761742. Epub 2020 May 25. PMID: 32397873; PMCID: PMC8204962.

108. Vázquez, Alexei. "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations." Physical Review E 67.5 (2003): 056104.

109. Giuraniuc, C. V., et al. "Trading interactions for topology in scale-free networks." Physical review letters 95.9 (2005): 098701.

110. 110.(Caetano-Anollés D, Caetano-Anollés K, Caetano-Anollés G. Evolution of macromolecular structure: a "double tale" of biological accretion and diversification. *Sci Prog*. 2018; 101:360-383.)

111. Caetano-Anollés G, Aziz MF, Mughal F, et al. Emergence of Hierarchical Modularity in Evolving Networks Uncovered by Phylogenomic Analysis. *Evolutionary Bioinformatics*. 2019;15. doi:10.1177/1176934319872980

112. Benson AR, Gleich DF, Leskovec J. Higher-order organization of complex networks. Science. 2016; 353:163-166.

113. Benson, Austin R., David F. Gleich, and Jure Leskovec. "Higher-order organization of complex networks." *Science* 353.6295 (2016): 163-166.

114. Michoel, Tom, et al. "Enrichment and aggregation of topological motifs are independent organizational principles of integrated interaction networks." *Molecular BioSystems* 7.10 (2011): 2769-2778.

115. Almaas, Eivind. "Biological impacts and context of network theory." Journal of Experimental Biology 210.9 (2007): 1548-1558.

116. A. Reményi, et al., The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network Mol. Cell, 20 (2005), pp. 951-962

117. J.P. Staley, J.L. Woolford Jr. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. Curr. Opin. Cell Biol, 21 (2009), pp. 109-118

118. Abbasian MH, et al., Global landscape of SARS-CoV-2 mutations and conserved regions. J Transl Med. 2023 Feb 25;21(1):152. doi: 10.1186/s12967-023-03996-w. PMID: 36841805; PMCID: PMC9958328.

119. Markov, P.V., et al. The evolution of SARS-CoV-2. Nat Rev Microbiol 21, 361–379 (2023). https://doi.org/10.1038/s41579-023-00878-2

120. Burley, S. K. et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Res. 49, D437–D451 (2021).

121. Burke, D.F., Bryant, P., Barrio-Hernandez, I. et al. Towards a structurally resolved human protein interaction network. Nat Struct Mol Biol 30, 216–225 (2023). https://doi.org/10.1038/s41594-022-00910-8

122. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science 373, 871–876 (2021).

123. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

124. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. Nat. Commun. 13, 1265 (2022).

125. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv https://doi.org/10.1101/2021.10.04.463034 (2021).

126. Humphreys, I. R. et al. Computed structures of core eukaryotic protein complexes. Science 374, eabm4805 (2021).

127. Modell, Ashley E., et al., "Systematic targeting of protein–protein interactions." *Trends in pharmacological sciences* 37.8 (2016): 702-713.)

128. Subramanian S, Kumar S. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics. 2004 Sep;168(1):373-81. doi: 10.1534/genetics.104.028944. PMID: 15454550; PMCID: PMC1448110.

129. Szaflik T, et al., An Analysis of ESR2 and CYP19A1 Gene Expression Levels in Women With Endometriosis. In Vivo. 2020 Jul-Aug;34(4):1765-1771. doi: 10.21873/invivo.11970. PMID: 32606145; PMCID: PMC7439897.)

130. Zeng C, et al., SARS-CoV-2 spreads through cell-to-cell transmission. Proc Natl Acad Sci U S A. 2022 Jan 4;119(1):e2111400119. doi: 10.1073/pnas.2111400119. PMID: 34937699; PMCID: PMC8740724. https://www.pnas.org/doi/full/10.1073/pnas.2111400119