Article

Not peer-reviewed version

A tiny viral protein, SARS-CoV-2-ORF7b: Functional molecular mechanisms.

Gelsomina Mansueto , Giovanna Fusco , Giovanni Colonna *

Posted Date: 8 March 2024

doi: 10.20944/preprints202403.0473.v1

Keywords: SARS-CoV-2-ORF7b, COVID-19, Interactomics, Topological analysis, Cluster analysis, Co-regulation network, Transcription Factors, microRNA, SARS-CoV-2 inter-tissue diffusion, programmed death.

Article

A Tiny Viral Protein, SARS-CoV-2-ORF7b: Functional Molecular Mechanisms

Gelsomina Mansueto [1], Giovanna Fusco [2] and Giovanni Colonna [3,*]

[1] Dipartimento di Scienze Mediche e Chirurgiche Avanzate, Università della Campania, L. Vanvitelli 80138 - Naples, Italy; gelsomina.mansueto@unicampania.it

[2] Istituto Zooprofilattico Sperimentale del Mezzogiorno. 80055 Portici, Naples, Italy; giovanna.fusco@izsmportici.it

[3] Medical Informatics AOU, Università della Campania, L. Vanvitelli. 80138 - Naples, Italy

* Correspondence: giovanni.colonna@unicampania.it

Abstract: This study presents the interaction with human host metabolism of SARS-CoV-2 ORF7b protein (43 aa), using a Protein-Protein-Interaction Networks analysis. After pruning, we selected from BioGRID the 51 most significant proteins among 2,753 proven interactions and 1,708 interactors, specific to ORF7b. We used these proteins as functional seeds and we got a significant network of 551 nodes via STRING. We performed topological analysis and calculated topological distributions by Cytoscape. Seven high ranked proteins as hub and seven as bottleneck following a hub-and-spoke network-architectural-model were found. Through this interaction model, we identified significant GO-processes (5,057 terms in 15 categories) induced in human metabolism by ORF7b. High statistical significance processes of dysregulated molecular cell mechanisms by the action of ORF7b were discovered. We detected disease-related human proteins and their involvement in metabolic roles, how they relate in a distorted way to signaling and/or functional systems, in particular intra- and inter-cellular signaling systems and the molecular mechanisms that supervise to programmed cell death, with mechanisms similar to that of cancer metastasis diffusion. A cluster analysis showed 10 compact and significant functional clusters, where two of them overlap in a Giant-Connected-Components core of 206 total nodes. These two clusters contain most of the high-rank nodes. ORF7b mainly acts through these two clusters, inducing most of the metabolic dysregulation. We conducted a co-regulation, transcriptional analysis by hub and bottleneck proteins. This analysis allowed us to define the transcription factors and miRNAs that control the high-ranking proteins and the dysregulated processes, within the limits of the poor knowledge that these sectors still impose.

Keywords: SARS-CoV-2-ORF7b; COVID-19; interactomics; topological analysis; cluster analysis; co-regulation network; transcription factors; microRNA; SARS-CoV-2 inter-tissue diffusion; programmed death

1. Introduction

This study aims to show the effects of the ORF7b viral protein of SARS-CoV-2 on humans, using significant experimental virus-host molecular interactions from BioGRID. Studying protein-protein interactions that contain information and metabolic strategies used by both the virus and its host allows us to understand functional relationships. We performed the analysis after functional enrichment to amplify less represented biological functions. SARS-CoV-2 encodes its genetic information in a single-stranded RNA, and ribosomes translate it into thirty-one different proteins. Viral action occurs through interactions with single human proteins or with protein complexes. To implement an effective action, the total number of viral proteins must be adequate for that of humans. About 5,000 viral particles are present in a single human cell during the peak time of infection (the first 3-4 days), along with a concentration of about 150,000 proteins/cell necessary for effective action,

 doi:10.20944/preprints202403.0473.v1

as estimated by reliable sources [1,2]. Other estimates [3,4] suggest that in the human cell, there are on average between two and four billion proteins, represented by a few thousand different types [3–5], and the average lifespan of each molecule is often measurable in a few dozen minutes. All this implies that each viral protein should interact with a target that has a rather limited time window, but viral proteins also have the same problem because of their turnover rate. Therefore, only a perfect knowledge of human metabolism, deriving from a co-evolution of coronaviruses with humans and/or mammals, can generate proteins that are effective. ORF7b is one of the smallest proteins of the virus [7], an accessory protein of only 43 amino acids with a central alpha helical segment, but its function is still unknown [6,191]. In recent years, various laboratories dedicated their activities to the research, purification, and characterization of the physical complexes between ORF7b2 and human proteome proteins with different methods and technologies. BioGRID [8] has collected and cured these experimental results within the "BioGRID COVID-19 Coronavirus Curation Project". BioGRID curates proven protein interactions between virus and humans and curators have classified the proteins according to criteria of statistical reliability. They have identified 2,753 physical interactions and 1708 interactors for ORF7b (accessed in July 2023). Thus, BioGRID presents an interactome of considerable interpretative complexity for this protein [9].

2. Materials and Methods

2.1. BioGRID

It is the source of experimental interactions of SARS-CoV-2-ORF7b (as of July 2023). (https://thebiogrid.org/4383871/summary/severe-acute-respiratory-syndrome-coronavirus-2/orf7b.html).

The quantitative SAINT analysis 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.

STRING [152,153] (https://string-db.org/) is a database of known and predicted PPIs. The curated interactions are direct (physical) and indirect (functional) associations. The interactions came from different sources (genomic context, high-throughput experiments, co-expression, previous knowledge, etc.) which are channeled into seven independent channels. 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 >0.9 (highest confidence) with the information from all seven sources.

Protein enrichment 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. We have used a selected set of protein by BioGRID as functional seeds. 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.

2.2. CYTOSCAPE and Network Topology Analysis

Cytoscape [154,155] through Network Analyzer was used to analyze the topological parameters of networks. 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 neighbors 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

 doi:10.20944/preprints202403.0473.v1

significantly higher degree values than the others and were primarily in two central modules that were closely connected and compact.

CentiScaPe - Centralities for undirected, directed and weighted networks. Centiscape [156] 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, as well as the topological location of proteins in their specific cellular compartments.

2.3. Evaluation of the HUB-and-Spoke Model

Many properties of a scale-free network depend on the value of the degree exponent of the power-law, γ [157]. Therefore, it is interesting to establish how the network properties vary with γ. The estimation of the expected maximum degree (also known as the natural cut-off) for a scale-free network, which represents the expected size of the largest hub, is based on the following formula [158]:

$$K_{max} \sim K_{min} \mathcal{N}^{1/\gamma-1} \quad \text{(Eq. 1}$$

Where Kmax and Kmin are the expected maximum and minimum degree of a node, respectively. N is the system size, in terms of the number of nodes. Based on Eq. 1, when $\gamma<2$ (as in our case) the link acquisition rate of the largest hub is faster than the growth of the network in terms of the number of nodes it contains. In this scenario, the high-degree nodes are attractive. Here the dynamic is of the "winner takes all" type. This leads to a hub-and-spoke network of topology, where all nodes are within a short distance of each other. Our interactome has a gamma value of 1.81, which favors at least one large topological module (metabolic module). A topological module represents an area of the network densely packed with nodes and links wherever nodes have a larger tendency to be connected to the nodes of the same area instead of the nodes placed outside the zone itself.

2.4. Cluster Analysis

For the cluster analysis, we have used the K-Means Clustering method [159]. 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. Here, for K, which defines the number of pre-defined clusters, we have used the value of 10 after various manual attempts to search the most reliable clusters in terms of compactness, metabolic functionality, and p-value.

2.5. GO and KEGG Pathway Analyses

To better research and show the biological function of proteins, we performed GO analysis, which included biological process (BP), cellular component (CC) and molecular function (MF). When the P value was below 0.05, we considered the results had statistical significance.

2.6. Network Analyst -- Comprehensive Gene Expression Profiling via Network Visual Analytics: TFs and miRNAs

The Network Analyst [160,161] interprets gene lists in a network. It enables the analysis of results present in the network via a powerful online network visualization framework. In protein-protein network analyses, the system also involves the existing relationships between genes, proteins, miRNAs, and human transcription factors, creating a co-regulatory network that is very useful for understanding the mutual relationships between these biological actors.

Databases: Gene-miRNA interactions - miRTarBase v8.0 Comprehensive experimentally validated miRNA-gene interaction data collected from miRTarBase.

 doi:10.20944/preprints202403.0473.v1

TF-gene-interactions - ENCODETranscription factor and gene target data derived from the ENCODE ChIP-seq data. The BETA Minus algorithm is used to selecting only peak intensity signals <500 and predicted regulatory potential scores <1 from the ENCODE ChIP-seq data for TF-gene-interactions.

Signaling - SIGNOR 2.0. The data is based on data from the SIGnaling Network Open Resource.

RegNetwork: Regulatory Network Repository of Transcription Factor and microRNA Mediated Gene Regulations. RegNetwork is a data repository of five-type transcriptional and posttranscriptional regulatory relationships for human and mouse:

1. tF → TF
2. TF → gene
3. tF → miRNA
4. miRNA → TF
5. miRNA → gene

This repository integrates curated regulations and the potential regulations inferred based on the transcription factor binding sites. Transcription factor (TF) and microRNA (miRNA) function at the transcriptional and posttranscriptional levels. It will be valuable for studying gene regulatory systems by integrating the prior knowledge of the transcriptional regulations between TF and target genes, and the posttranscriptional regulations between miRNA and targets. The conservation knowledge of the transcription factor binding site (TFBS) can also be implemented to couple the potential regulatory relationships between regulators and their targets. From RegNetwork, we can query and identify the combinatorial and synergic regulatory relationships among TFs, miRNAs, and genes [162].

2.7. Protein Intrinsic Disorder and Secondary Structure Prediction

We have used two servers on line, Jpred 4 and IUPred2A. Jpred is a web server that takes protein sequences, and from these predicts the location of secondary structures using a neural network called Jnet. They show the prediction as a graph. IUPred2A **[163,164]** is a combined web interface that allows to identify disordered protein regions using IUPred2 and disordered binding regions using ANCHOR2. IUPred2A can identify disordered protein regions by analyzing their sequence, regardless of whether they are stable. Upon visually inspecting the graphic outputs of both predictive systems, we quickly identified disordered segments in most of the examined proteins, whether viral or human. These results were not displayed because they required a large space.

2.8. 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 F.D.R. =< 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.

 doi:10.20944/preprints202403.0473.v1

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.

2.10. 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 due to genetic enrichment, indirectly indicating increased gene expression, while the smaller the p value, the greater the certainty that event will occur. 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. This product will be greater when "Strength" has a very high value and p has a slight value [the most favorable situation for evaluating an effect positively is that represented by the extremes of their numerical values, very high and slight, respectively]. This facilitates comparisons and evaluation of 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 statistically 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 functionally 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.

3. Results

3.1. Source of the Data

Fundamental experimental data supporting the role of SARS-CoV-2 in human infection are accumulating. BioGRID, one of the most important biomedical interaction repository, compiled comprehensive datasets of all physical interactions between the proteins of SARS-CoV-2 and the human proteome through the BioGRID COVID-19 Coronavirus Curation Project [8,10]. Curators selected interaction data caming from purification processes where researchers used physical methods such as Affinity Capture-MS and Proximity Label-MS. Interactions and their molecular interactors were classified into various levels of significance. With the protein ORF7b (P0DTD8 - NS7B_SARS2, UniProt), BioGRID classified 1,708 unique curated physical interactors [11–17], involved in 2,753 interactions (accessed in July 2023). They are unique in being non-redundant and having high confidence interactions at high throughput, associated with high score values of statistical filtering, as determined by using SAINT (Significance Analysis of INTeractome) express version 3.6.0 [11–17].

 doi:10.20944/preprints202403.0473.v1

3.2. The Representation of ORF7b Data Using Interactomes

Figure 1 shows the circular network of human ORF7b-interacting proteins calculated by BioGRID. Since not all physical interactions flow into a real biological function, the concentric representation of the nodes shows different levels of reliability. Therefore, we used the densest layers as functional seeds. The nodes selected in this study have proven physical interaction through at least two different physical methods. The interaction should be non-redundant and high-throughput with optimal statistical significance between BioGRID levels 6 and 4. These options allowed us to select nodes with curated unique interactions.

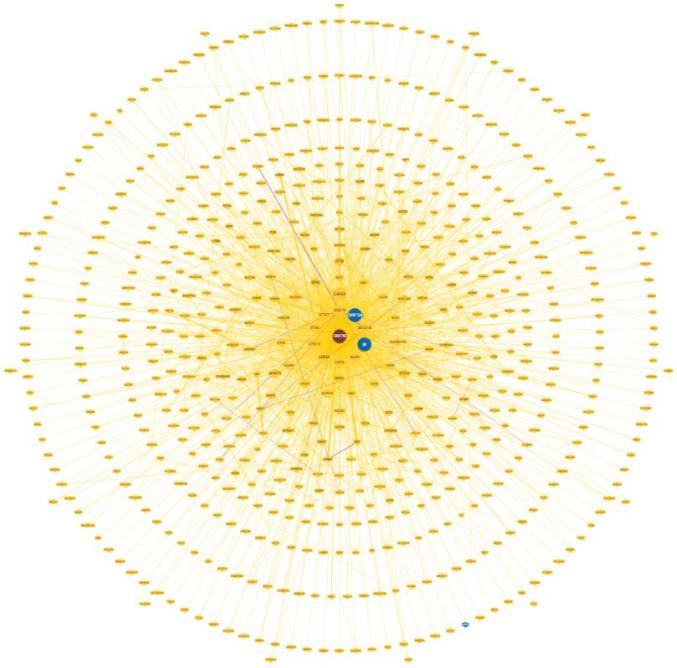


Figure 1. – Circular network of SARS-CoV-2-ORF7b and human host PPI (from BioGRID). Circles within circles representation show the layers closest to the center as more highly connected. BioGRID also suggests the likelihood of direct/indirect interaction of ORF7b (in dark red) with other viral proteins (ORF3a and M, in blue). The proteins used in the present analysis are among those of the most densely represented central area.

In **Figure 1S (**Supplements**)**, we show an ARBOR representation of the network calculated by BioGRID with a minimum evidence value of 4, which illustrates the level/association relationships very well. An interactome shows the one-to-one mapping of all interactions, which turns the interactome into an information system [18]. The goal is to decode the functional information of this biological map, the macroscopic properties of which are unpredictable and emergent properties of the system [19,20]. Its inherent complexity makes it difficult, if not impossible, to decode individual hidden molecular information. The datasets curated by BioGRID for each SARS-CoV-2 protein represent a suitable starting material. The list of 75 ORF7b interactors with significant levels ranging from 6 to 4 is available in **Table S1.** Through the STRING platform [21] we calculated the corresponding interactome (**figure 2S** in Supplements) with a score of 0.9 and with all 7 data source channels active, to gain as much information as possible. But the graph shows 54 proteins (72%) unconnected. So, we added 500 first order proteins to enrich the interactome and increase the functional relationships (**figure 3S** in Supplements). In this new graph, we also had to eliminate some parental proteins that were still disconnected, leaving 51 final parental proteins that were the basis of our enriched interactome. Network pruning helps eliminate artifacts due to noisy information [22] while enrichment helps amplify those biological processes that are difficult to define because of their poor representation. **Figure 2** shows the interactome got after pruning and enrichment. The interactome now appears compact, with all nodes connected.

 doi:10.20944/preprints202403.0473.v1

Typically, proteins that share similar functional information should appear as a compact set of nodes and edges (sub-graphs) performing one or more macroscopic functions. Subgraphs contain molecular partners that have relational links and perform similar functional activities. Analyzes of metabolic processes with Gene Ontology or KEGG allow us to evaluate the increase in functional annotations.

Many and rather compact peripheral modules with a large and very compact central module characterise this interactome. The peripheral modules suggest functional protein complexes. For example, the module at the top of the figure is very rich in ribosomal subunits and, very close to it, many proteins belonging to the translocon complex can be identified. While the complex on the right is rich in ATPase subunits characteristic of the proton-transporting vacuolar protein pump (V)-ATPase, required for acidification of secretory vesicles. These complexes represent the set of metabolic machinery necessary for normal cellular life. Surprisingly, the large central component shows nodes intra-connected, representing a significant fraction (37%) of the network's nodes. Components with these characteristics are called Giant Connected Components (GCC) [23]. This type of component is often present in scale-free networks of which it is an important substructure. GCCs control the topological growth of the network, and so its evolution [24]. Its capacity to aggregate new nodes and functions makes it a very compact system with a notable increase in the interaction turnover rate of new proteins [24].

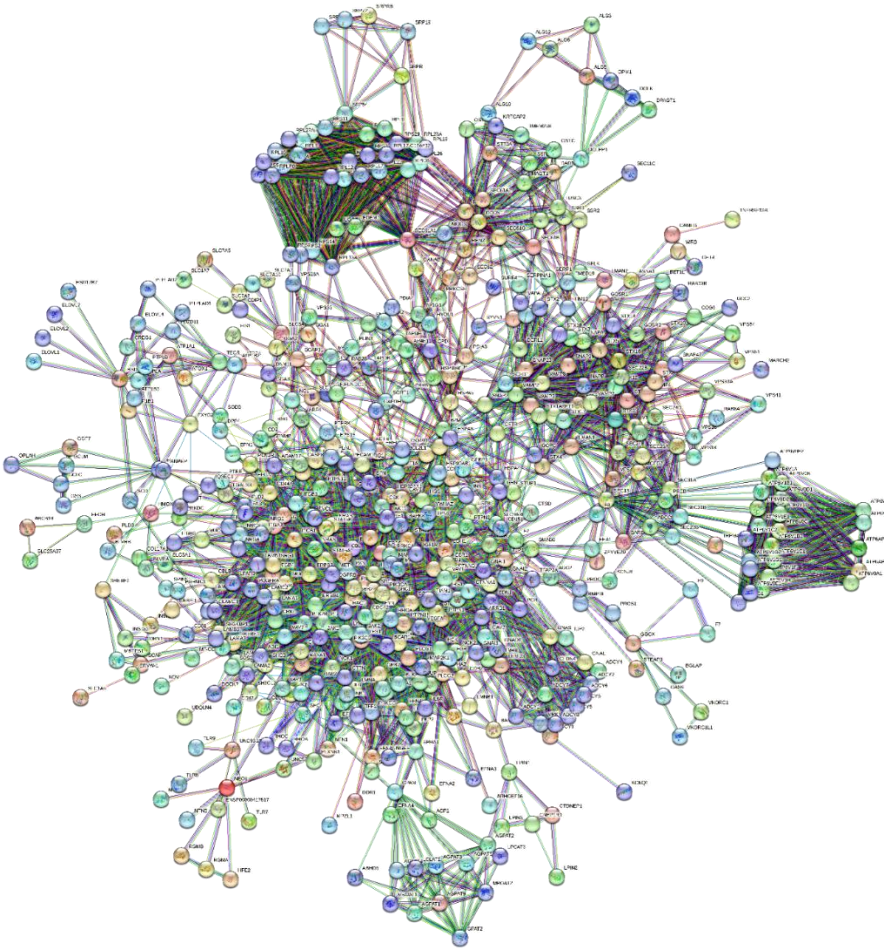


Figure 2. Interactome of 51 human proteins functionally involved with ORF7b2, enriched with 500 first order proteins. An overall look reveals the involvement of peripheral compact groups of nodes that can represent specific functional modules or even particular protein complexes. Network calculated by STRING and the score is 0.9. The number of edges we have is greater than the number of nodes in a similar random network we can calculate (PPI enrichment p-value < 1.06e-16). We show the topological parameters in **Table 1**.

 doi:10.20944/preprints202403.0473.v1

We can find a demonstration of this compactness in **Table S2** and F**igure 4S** (in Supplements). The figure shows the distribution graph of the mean shortest paths as a function of the degree of the single nodes. The 30 nodes with the highest ranks, i.e., with the greatest connectivity in the network, are those with the lowest average shortest path-length. These nodes are all concentrated in the GCC. Thus, this network has a "giant component", where almost every node is easily reachable from almost every other node in GCC, through a dense net of interactions. New nodes will massively join the GCC in a non-linear and unpredictable way to create biological functions, as GCC is a set of functionally very attractive metabolic nodes. This helps create the set of functions of this metabolic module [24]. Typically, as the network grows, the giant component will continue to incorporate a significant fraction of incoming nodes. This means that we should find the main and crucial functional activities integrated into this subgraph.

3.3. Principal Characteristics of the Interactome

We transferred the interactome to Cytoscape [25] and analyzed it with the help of CentiScaPe (v2.2), Analyze Network [26], and STRING-app [27,28], which generated a Table of Nodes containing various columns with the quantitative values of many topological and functional parameters. This allowed the evaluation of characteristic topological and functional features for each node of the interactome.

The value of parameters in **Table 1** tells us we are considering a network made up of many independent and compact peripheral modules, which exchange relationships with fewer connections between them, albeit essential. The large diameter, network heterogeneity, and low-density support this view [29]. The diameter also suggests components quite distant from the central module. While the shortest average path length, which gives the distance between two connected nodes, is a metabolic advantage because small average lengths minimize transition rates between metabolic states in response to external stress. The clustering coefficient also supports this topology. It is a basic index for local density in a network and is a measure of the degree to which nodes in a graph group together. It takes values $0 \leq C \leq 1$, thus a value of 0.549 shows a tendency to form clusters, where each node shows an average of 16.817 neighbors. This coefficient of aggregation, according to Barabasi [30], decreases with the increase in nodes.

Table 1.

Summary Statistics of network*		**Notes**
Number of nodes	551	
Number of edges	4648	**
Avg. Number of neighbors	16,871	Average connectivity of the nodes
Network diameter	9	
Characteristic path length	3.666	
Clustering coefficient	0.549	$0 \leq C \leq 1$
Network density	0.031	
Network heterogeneity	1.057	Tendency to contain hub nodes
Network centralization	0.259	The extent to which certain nodes are far more central than others
Connected Component	1	***

*) Calculated by Cytoscape Network Analyzer, which computes a comprehensive set of topological parameters [25, 26]. **) Most nodes (77%) with a score of 0.9 have a very large component of the scientific information necessary to calculate the interactions that derive from the Text Mining channel with only a partial presence of data coming from the Experiments channel. While, only 15.7% of the edges show a full score of 0.9, deriving from the "Experiments" channel alone, proving that their interactions are experimental. ***) This value is "1" to show that all nodes in the network are connected to each other. The presence of unconnected components (CC >1) alters the calculations of the topological parameters, making them unreliable. This is the fundamental reason for pruning. A single component accounts for strong network connectivity. Calculation by Cytoscape.

 doi:10.20944/preprints202403.0473.v1

The **Figure 3** shows the characteristic power distribution of nodes of a scale-free network, where the vast majority of nodes have very few connections, and only few (HUBs) have a very large number of connections [31]. This distribution is a defining characteristic of the biological network regardless of the experimental approach [32] and is important in understanding the system's behavior. The power law exponent highlights a configuration for scale-free networks that minimize the number of nodes needed to control the entire system [33,34]. In the figure, we highlight the seven HUB nodes (EGFR, SRC, PIK3R1, PIK3CA, GRB2, and HRAS), which have superior ranks compared to all others, also remembering that the GCC includes the top 30 nodes with the highest ranks. Hub nodes model the architecture of metabolic modules and EGFR, which serves multiple critical functional roles in the cell, is the highest degree interactomic hub node also because of its exceptional capacity for PTMs (see **figures 5S).**

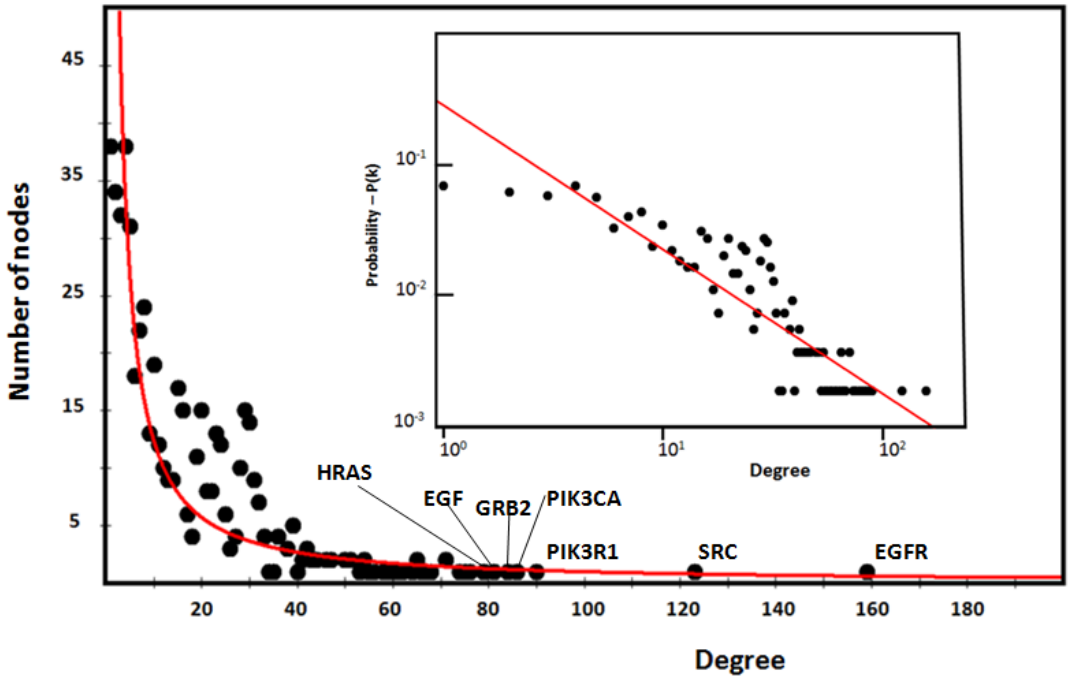


Figure 3. Node Distribution. The distribution follows a free scale distribution based on the power law. In the inset, we present the same nodes on a log-log scale, with the best fit of data shown in red. The function used for the fit is $f(x) = a*x^{b,}$ where the values of a, b, and R^2 are 0.29, -1.89, and 0.62, respectively. A significant p-value of 1.0exp-16 of the interactome analysis and a good correlation index underscore a strong expectation of preferential relations or associations among nodes following their enrichment.

We need alternative tests to prove the accuracy of our observations and hypotheses and to decode the information due to the actual functional activities in which ORF7b2 is involved. The following tables will show the most significant contents of some important functional categories. To evaluate the importance of each functional property, we will use the p-value as the evaluation criterion [5] for the main significant processes. STRING calculated the tables with the methods and techniques of GO analysis.

3.4. Quantitative Evaluation of the Biological Functionalities in the Interactome

Table 2 shows the overall picture of the many functional activities performed by the entire network. Over 10,000 significant PubMed publications were used to provide coherent information on the 5,057 functional terms. STRING calculated the entire interactome using this knowledge base. This assures us that the functional relationships taken into consideration are very robust and that the pruning operation reflected real knowledge gaps in the considered node properties. The spectrum of

 doi:10.20944/preprints202403.0473.v1

biological activities induced by ORF7b2 appears remarkably broad in 15 categories and, therefore, both difficult to define and to study thoroughly. We have evaluated and selected the functional activities from time to time, as each of the 5,057 terms reported in **Table 2** has a statistical value (p-value) that is always less than 0.05, ensuring their significance. In this study, we will try to give a comprehensive view of the metabolic and molecular activity induced by ORF7b. Future studies will try to go into more detail.

Table 2.

Action	Enriched terms
Biological Process (Gene Ontology):	*1690 GO-terms*
Molecular Function (Gene Ontology):	*166 GO-terms*
Cellular Component (Gene Ontology):	*267 GO-terms*
Reference publications (PubMed):	***>10,000 publications***
Local network cluster (STRING):	*137 clusters*
KEGG Pathways:	*195 pathways*
Reactome Pathways:	*494 pathways*
WikiPathways:	*259 pathways*
Disease-gene associations (DISEASES):	*112 diseases*
Tissue expression (TISSUES):	*186 tissues*
Subcellular localization (compartments):	*249 compartments significantly*
Human Phenotype (Monarch):	*1002 phenotypes*
Annotated Keywords (UniProt):	*99 keywords*
Protein Domains (Pfam):	*63 domains*
Protein Domains and Features (InterPro):	*118 domains*
Protein Domains (SMART):	*20 domains*
All enriched terms (without PubMed):	***5,057 enriched terms in 15 categories***

Table 3 shows the most significant biological functions (GO-Biological Processes) among the 1,690 related to the human proteome following the action of ORF7b. The principal activities involve the control of intracellular transport, also by vesicles, and the control of their localization in the cell. The set of cellular processes includes the transportation, binding, and holding of a protein complex or organelle in a specific position. A transporter or group of transporters facilitates the directed movement of molecules or cellular complexes into or out of a cell, or between cells, to effect transmembrane, microtubule-based, or vesicle-mediated transport. A significant value ranging from a p-value of 1.0e-77 down to 0.05 marks all 1,690 activities. Enzymes and signaling pathway receptors also appear to be possible prime targets, also considering the large number of human proteins involved. In particular, the series of molecular signals started with an extracellular ligand binding to a receptor with tyrosine kinase activity on the surface of the target cell and ending with regulating a downstream cellular process. The statistical significance of these biological actions is very high, as is the number of proteins involved. However, the table shows a comprehensive picture of 1,650 functional activities that belong to both the virus and the cell in performing their respective strategies of attack or defense. A part of these activities also refers to the basal metabolic activities for the maintenance of normal vital functions (housekeeping functions). As we will see later, it will be possible to extract the specific activities of the virus.

Table 3. Biological Functions.

GO Term ID	Term description	Number of involved proteins	p-value
GO:0051179	Localization	378	2.01e-77
GO:0006810	Transport	320	3.04e-67
GO:0007169	Transmembrane receptor protein tyrosine kinase signaling pathway	124	1.23e-66
GO:0051234	Establishment of localization	322	7.72e-66
GO:0015833	Peptide transport	187	1.09e-62
GO:0051649	Establishment of localization in cell	230	3.37e-62
GO:0051641	Cellular localization	254	1.29e-60
GO:0015031	Protein transport	181	7.86e-60

 doi:10.20944/preprints202403.0473.v1

GO:0007167	Enzyme linked receptor protein signaling pathway	131	7.95e-59
GO:0008104	Protein localization	213	1.46e-58
GO:0045184	Establishment of protein localization	183	2.85e-58
GO:0016192	Vesicle-mediated transport	189	1.18e-53
GO:0032879	Regulation of localization	229	2.95e-51
GO:0009987	Cellular process	546	4.49e-51
GO:0046907	Intracellular transport	168	1.19e-49

The Table 4 depicts the location in the cell where the most statistically significant functional activities (as presented **in Table 3)** occur**.** Many cell membranes, cytoplasm, as well as protein complexes, are metabolically involved. Of particular interest is the significant activity performed by the SNARE complex, specifically involved in driving vesicles and endosomes towards the correct cellular target, also providing for the correct docking. SNARE proteins (**SNA**p **RE**ceptor, i.e., Soluble *N*-ethylmaleimide-Sensitive Factor Attachment Proteins) are a family of cytosolic proteins involved in vesicular fusion with the target membrane during intracellular transport and exocytosis [35]. SNAPs interact with proteins of the SNARE complex during the recycling of the fusion complex components [36]. We know that interference with the function of SNAP proteins is associated with many pathological processes, such as colorectal cancer [37], epilepsy [38] or Huntington's disease [39]. However, it is the post-translational process by which a PTM protein (a proteoform) trans-locates from the ER to its final destination, which drives function. This process also includes tethering and docking steps that prepare vesicles for fusion.

Table 4. CELLULAR LOCALIZATION OF BIOLOGICAL FUNCTIONS.

GO Term ID	COMPARTMENT	Number of involved proteins	p-value
GOCC:0016020	Membrane	399	2.58e-92
GOCC:0012505	Endomembrane system	302	1.36e-91
GOCC:0031090	Organelle membrane	243	2.07e-77
GOCC:0098796	Membrane protein complex	189	7.35e-74
GOCC:0005737	Cytoplasm	437	1.41e-73
GOCC:0031982	Vesicle	213	5.13e-62
GOCC:0098588	Bounding membrane of organelle	174	3.17e-58
GOCC:0005783	Endoplasmic reticulum	133	6.29e-55
GOCC:0098805	Whole membrane	156	1.76e-53
GOCC:0110165	Cellular anatomical entity	531	2.59e-51
GOCC:0005789	Endoplasmic reticulum membrane	105	4.93e-51
GOCC:0042175	Nuclear outer membrane-ER membrane network	106	1.79e-50
GOCC:0031410	Cytoplasmic vesicle	177	1.29e-49
GOCC:0032991	Protein-containing complex	306	4.76e-44
GOCC:0043226	Organelle	437	1.20e-41
GOCC:0043227	Membrane-bounded organelle	406	5.80e-41
GOCC:0005622	Intracellular	462	8.33e-38
GOCC:0043229	Intracellular organelle	407	4.82e-34
GOCC:0005829	Cytosol	201	2.18e-32
GOCC:0005886	Plasma membrane	220	3.75e-30
GOCC:0031201	SNARE complex	34	3.79e-30
GOCC:0043231	Intracellular membrane-bounded organelle	349	6.22e-30

Table 5 (Reactome) shows the most statistically significant molecular mechanisms in which ORF7b might involve the human proteome. It contains biomolecules that perform precise metabolic and signaling activities and their relationships organized into biological pathways. Beyond the various interferences on important metabolic pathways, it is interesting to note the metabolic functions shown, such as, Nervous system development, Immune System, Infectious disease, Hemostasis, Innate Immune System, Platelet activation, Insulin receptor Signaling, Viral mRNA Translation and Cell-Cell communications. Although they are normal vital metabolic functions with

 doi:10.20944/preprints202403.0473.v1

high statistical significance, the parallelism with the known clinical effects of COVID-19 on the human organism [40,41] should not be overlooked, which is surprising.

Table 5. REACTOME.

Term ID	Molecular Mechanism	Number of involved proteins	p-value
HSA-9006934	Signaling by Receptor Tyrosine Kinases	140	4.44e-84
HSA-1643685	Disease	189	2.66e-63
HSA-422475	Axon guidance	101	6.79e-45
HSA-9675108	Nervous system development	103	6.79e-45
HSA-168256	Immune System	176	6.06e-41
HSA-5663205	Infectious disease	115	6.30e-41
HSA-162582	Signal Transduction	204	2.84e-37
HSA-5653656	Vesicle-mediated transport	95	2.70e-34
HSA-199991	Membrane Trafficking	92	5.34e-34
HSA-392499	Metabolism of proteins	163	3.25e-33
HSA-109582	Hemostasis	89	1.02e-32
HSA-1799339	SRP-dependent cotranslational protein targeting to membrane	45	2.19e-31
HSA-168249	Innate Immune System	111	1.33e-30
HSA-1227986	Signaling by ERBB2	35	2.43e-30
HSA-74752	Signaling by Insulin receptor	38	5.77e-29
HSA-177929	Signaling by EGFR	33	5.40e-28
HSA-4420097	VEGFA-VEGFR2 Pathway	39	5.35e-27
HSA-202733	Cell surface interactions at the vascular wall	42	3.19e-25
HSA-76002	Platelet activation, signaling and aggregation	52	4.89e-24
HSA-6811558	PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling	37	5.16e-24
HSA-5683057	MAPK family signaling cascades	52	1.19e-20
HSA-5684996	MAPK1/MAPK3 signaling	49	1.37e-20
HSA-77387	Insulin receptor recycling	21	1.07e-19
HSA-192823	Viral mRNA Translation	27	1.54e-16
HSA-1500931	Cell-Cell communication	30	1.91e-15

The spectrum of possible viral interference also might involve intracellular transport mechanisms and cell-cell communications. Many of these "actions" have a deep impact on human biology and inter-organ signaling, according to recent research on the effects of COVID-19 on the human organism [42,43]. In particular, we relate the most significant one to signaling by receptor tyrosine kinases (RTKs), a family of proteins that act as cell surface receptors for various factors, such as cytokines and hormones. These receptors control many cellular processes but have also a crucial role in the development and progression of many types of cancer [44,45]. It is also interesting to highlight the high significance in this interactome of some activities, such as "Cell surface interactions on the vascular wall", "Platelet activation", "Insulin receptor recycling", "Viral mRNA translation", "Cell-cell communication".

By using proteins directly involved with ORF7b, we extracted relevant activities in this interactome selectively from the human proteome. The symptoms in COVID patients, including thrombophilic alterations [46], hyperglycemia [47], and systemic spread of infected cells [48], may not be independent, as their underlying mechanisms, as found in Reactome, all appear to have the involvement of ORF7b, which may be the underlying cause.

The number of human tissues and organs that are potential targets of ORF7b is also staggering. **Table 6** shows these tissues/organs, which are important constituent of human body through many cell types.

 doi:10.20944/preprints202403.0473.v1

Table 6. HUMAN TISSUES INVOLVED WITH ORF7b.

TERM ID	HUMAN TISSUES INVOLVED WITH ORF7b	Number of involved proteins	p-value
BTO:0000345	Digestive gland	233	4.73e-56
BTO:0001491	Viscus	322	1.59e-54
BTO:0001489	Whole body	504	2.18e-45
BTO:0000522	Gland	356	2.76e-45
BTO:0000759	Liver	178	2.17e-44
BTO:0001488	Endocrine gland	323	1.15e-37
BTO:0003091	Urogenital system	341	3.03e-36
BTO:0000227	Central nervous system	303	1.41e-35
BTO:0001484	Nervous system	307	3.23e-35
BTO:0000449	Fetus	125	1.68e-32
BTO:0001078	Placenta	119	1.40e-30
BTO:0000081	Reproductive system	308	5.01e-30
BTO:0003099	Internal female genital organ	183	5.01e-30
BTO:0000174	Embryonic structure	159	7.75e-28
BTO:0000203	Respiratory system	127	9.81e-28
BTO:0000083	Female reproductive system	292	1.71e-27
BTO:0000089	Blood	136	1.39e-26
BTO:0000570	Hematopoietic system	172	6.39e-26
BTO:0000763	Lung	105	3.59e-23
BTO:0000988	Pancreas	72	6.06e-23
BTO:0000431	Excretory gland	106	8.44e-21
BTO:0003092	Urinary system	97	3.43e-19
BTO:0001244	Urinary tract	97	4.17e-19
BTO:0000671	Kidney	86	5.49e-19
BTO:0001129	Prostate gland	58	8.08e-19
BTO:0000132	Blood platelet	50	2.81e-18
BTO:0000511	Gastrointestinal tract	116	4.50e-17
BTO:0000131	Blood plasma	51	1.46e-16
BTO:0000574	Hematopoietic cell	77	1.21e-14
BTO:0000082	Male reproductive system	148	3.21e-14
BTO:0000751	Leukocyte	72	3.31e-14
BTO:0000080	Male reproductive gland	138	6.39e-13
BTO:0000254	Female reproductive gland	145	7.61e-12
BTO:0005810	Immune system	96	3.68e-11
BTO:0003096	Internal male genital organ	122	4.92e-11
BTO:0000088	Cardiovascular system	70	4.40e-10
BTO:0000421	Connective tissue	63	1.19e-09
BTO:0000439	Eye	59	1.33e-09
BTO:0000706	Large intestine	54	1.57e-09
BTO:0000202	Sense organ	69	1.98e-09
BTO:0000855	Lymph	25	4.56e-09
BTO:0001085	Vascular system	38	9.90e-09
BTO:0001424	Uterus	67	1.11e-08
BTO:0000269	Colon	46	3.05e-08
BTO:0001363	Testis	85	2.48e-05

These tissues/organs share many of the previously described metabolic activities to varying degrees. Therefore, even if not all, they are potential targets of the virus where it finds the optimal metabolic conditions for its replication [49]. The need to expand the list of terms in this table arises from the need to show the many target tissues of the virus with a significant potential. It is amazing how a tiny protein like ORF7b could induce so wide effect. This also means that the protein appears to be an authoritative candidate for altering the molecular mechanisms that keep cells in contact with each other [50–52]. Dysregulating these mechanisms might free the cells to spread without a programmed death [53,54].

This **TABLE** shows a long list of the various organs in the abdominal cavity, which are potential targets of the action of this protein, and validates the clinical observations that covid is a systemic disease. The high statistical values suggest the enormous potential of the strategy implemented by

 doi:10.20944/preprints202403.0473.v1

SARS-CoV-2 in hitting the human body. Some objectives are of particular interest. Nervous system (central and peripheral), human reproductive system (male and female), placenta and fetus, blood and hematopoietic system should alert us to the consequences encountered in the long-covid. Long-covid is showing symptoms that suggest the involvement of these specific organs and tissues as well.

An important index is also the high total number of proteins involved in each of the multiple functional activities represented in the tables previously reported. Considering the finite number of proteins in the interactome and the large number of them involved in many and different metabolic activities, this suggests that there is a high probability that single proteins may be involved in numerous functions. But all this also suggests that, in the event of a viral infection, a single human protein can perform many functional activities, some for the benefit of the cell and others for the benefit of the virus. KEGG pathways can infer higher-level functions and metabolic utilities of the human system from genomic and proteomic data. It groups genes and/or proteins into "pathways" as lists of genes/proteins taking part in the same metabolic process. Thus, KEGG is very useful for computational analyzes, including metabolic modeling and simulation according to systems biology, and translational research in disease development. KEGG's results show a wide range of activity. The breadth and diversity of the responses (195 pathways) and their statistical significance would require more space to highlight many of them. However, we have included the most probable **in Table 7**. These pathways reflect precise connections with the functions reported in the previous tables, identifying and endorsing their metabolic role. We can only identify the most significantly represented functions, but we cannot at this stage establish a direct correlation to viral activity.

Table 7.

Pathway	Description	Number of involved proteins	p-value
hsa04012	ErbB signaling pathway	50	3.02e-41
hsa04510	Focal adhesion	64	2.27e-40
hsa01521	EGFR tyrosine kinase inhibitor resistance	46	4.29e-38
hsa04151	PI3K-Akt signaling pathway	74	1.16e-36
hsa04141	Protein processing in ER	55	1.44e-35
hsa04015	Rap1 signaling pathway	51	3.72e-28
hsa04014	Ras signaling pathway	52	3.76e-27
hsa05206	MicroRNAs in cancer	45	1.48e-26
hsa04935	Growth hormone synthesis, secretion action	40	3.80e-26
hsa04130	SNARE interactions in vesicular transport	27	1.38e-25
hsa04062	Chemokine signaling pathway	45	2.35e-24
hsa04145	Phagosome	40	9.33e-24
hsa04360	Axon guidance	43	2.15e-23
hsa04072	Phospholipase D signaling pathway	39	2.11e-22
hsa04917	Prolactin signaling pathway	30	2.84e-22
hsa04150	mTOR signaling pathway	39	4.38e-22
hsa04810	Regulation of actin cytoskeleton	42	3.07e-20
hsa01522	Endocrine resistance	31	4.10e-20
hsa04915	Estrogen signaling pathway	35	4.10e-20
hsa04722	Neurotrophin signaling pathway	33	4.29e-20
hsa04919	Thyroid hormone signaling pathway	32	9.72e-19
hsa04664	Fc epsilon RI signaling pathway	26	1.16e-18
hsa04010	MAPK signaling pathway	45	5.16e-18
hsa04721	Synaptic vesicle cycle	26	1.05e-17
hsa04660	T cell receptor signaling pathway	29	1.05e-17
hsa04662	B cell receptor signaling pathway	26	2.90e-17
hsa04650	Natural killer cell mediated cytotoxicity	30	7.46e-17

So far, we have examined the spectrum of functional/molecular activities present in an infected cell and, in particular, those involved by ORF7b. Once we have defined the principal functions, we need to highlight which single proteins favor the virus by "playing a double game".

 doi:10.20944/preprints202403.0473.v1

3.5. Exploring the Physical Basis of Cytoskeletal Alterations Caused by ORF7b

The propagation of a virus to uninfected cells makes up a crucial phase in its life cycle, achieved through the liberation of novel viral particles from the infected cell. The ability of ORF7b to induce changes in the cytoskeleton that could promote the spread of infected cells is not coincidental. As we have previously discussed, these changes seem to derive from dysregulations induced at the cytoskeleton level. These results, however, suggest different biological events from those already known, not only a spread of viral particles after cell rupture but also a spread of entire infected cells to distant tissues, exactly like tumor metastases. Therefore, this aspect needs a greater attention. The key processes for modifications of the cell membrane, or that of cellular compartments, should pass through direct deformations caused by specific proteins that interact with the membrane [165], or even through indirect deformation by the cytoskeletal structures [166]. Therefore, the cytoskeleton is one of the key driving forces, with a close association to these events [167].

Unfortunately, understanding the influence of these molecular processes on the physical structure of the membrane is still an unsolved challenge, despite a slight improvement in our understanding of the underlying physical basis. Until now, it has been difficult to quantify the forces present in living cells within these processes. However, we now have a first, albeit crude, quantitative understanding of force production and distribution at the molecular level using clathrin-mediated endocytosis as a model [165,166]. During endocytosis, the actin cytoskeleton generates forces that are transmitted to the plasma membrane through a multi-protein coat, leading to membrane deformation. Although the exact extent of these forces remains uncertain, we can highlight a phenomenon of accumulation and redistribution of force within the endocytic mechanism. This has led to the widespread belief that the EPNs and Hip1R proteins transmit the force generated by the assembly of the actin to the plasma membrane [168,169]. As both protein types also attach to clathrin and other coat proteins, it is plausible that the transmission of forces to the membrane might occur through multiple pathways [170,171].

However, we know which eukaryotic genes/proteins actively engage in these processes, serving as either components or regulators of the cytoskeleton, while an intricate interplay between lipids and proteins controls the membrane remodeling during intracellular trafficking [172]. Noteworthy examples include MTOR, CTNNA1 (alpha 1 catenin), CTTN (cortactin), ITGBs (integrins), CDH1, CDH2 (cadherins), ACTB (actin B), and EPNs (Epsin family). A check of the interactome in **Figure 2** identifies all eight proteins and various members of their families (please also refer to the accompanying excel file for the comprehensive list and node degrees). This observation drew our attention to the intriguing possibility regarding the potential involvement of specific human proteins, in particular those associated with cytoskeletal modifications and negative regulation processes, in the mechanism of SARS-CoV-2 spread to non-infected cells and tissues. We used these proteins as seeds to tease out their functional relationships within the human proteome. **Figure 4** illustrates the specific and close relationships between them during their involvement in the processes that impact the organization of the cytoskeleton. Using a specific feature of STRING, the proteins involved in the same biological process were highlighted and colored (see Methods).

The network comprises all the human proteins involved in cytoskeleton dynamics. Since they are all reported in BioGRID as actively interacting, this suggests direct physical and/or functional associations. Among those of high rank, some, such as ACTB, are involved in a single dysregulated process (one color), others, such as MTOR, are involved in the management of multiple dysregulated processes (various colors). However, these interactions imply that SARS-CoV-2 exploits the host cell's proteins involved in processes regulated by CDH1, CDH2, EPN1, EPN2, CTNNA1, ITGB1, MTOR, ACTB, CTNNA1, and CTTN. This certainly affects cellular functions related to cell adhesion, signaling pathways, cytoskeletal organization, and programmed death through a Viral Hijacking of Cellular Machinery. But, these specific interactions also suggest potential roles for these cellular proteins in stages of the viral life cycle. In fact, their presence shows that these host proteins contribute to SARS-CoV-2 infection dynamics and pathogenesis, thus becoming appropriate therapeutic targets. However, further observations are important. Structural models of protein interfaces and the potential impact of post-translational modifications are crucial to understanding molecular

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0473.v1

mechanisms based on interactions because alteration of these characteristics might change protein-protein interactions and related biological functions. Many of the cytoskeletal proteins possess disordered structural domains and many phosphorylation sites. MTOR, serine/threonine protein kinase, in presence of RPTOR (Regulatory-associated protein of mTOR) and RICTOR (RICTR, Rapamycin-insensitive companion of mTOR) and through mTORC1 and 2 complexes, directly or indirectly controls the phosphorylation of at least 800 proteins and actin cytoskeleton is specifically MTOR sensitive [173–175]. DEPTOR (DEP Domain Containing MTOR Interacting Protein) is a negative regulator of TOR signaling and of mTORC1 and 2 pathways, inhibiting activity of both complexes [176,177]. This leads to negative regulation of cell size, and negative regulation of protein kinase activity. MTOR, DEPTOR, RICTOR, and RPTOR are all part of the interactome and communicate extensively. Thus, the relationships between them validate the various dysregulations in **Figure 4** and **Table 8**. A last consideration is that another viral protein also interacts with the cytoskeleton, it is the N protein, which plays various roles in the life cycle of the coronavirus [178]. Here we want to underline that the N protein physically interacts with ACTB [179], reconfiguring and manipulating the cytoskeleton as also happens for other viruses. This protein, as we will see in **Table 10**, also physically interacts with ORF7b. The N protein was mentioned because it is the SARS-CoV-2 protein that is involved in the formation of liquid droplets (see in "Discussion" for more details), a little discussed issue in the infection of this virus.

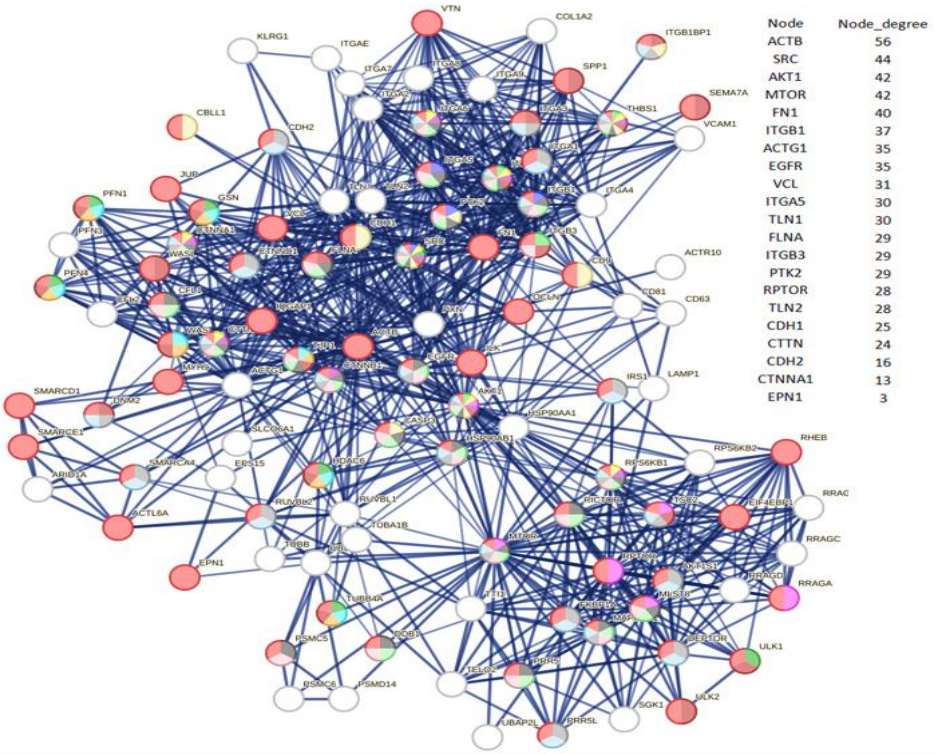


Figure 4. – Relationships among cytoskeleton related proteins. Network (top left side) - Score 0.7 (high confidence); all seven source channels are active; enrichment of the 8 basic proteins as functional seeds with 100 first-order proteins. Enrichment up to 100 proteins was necessary to achieve integration of all eight proteins into the network without expanding the number of functions too much. Topological data: number of nodes, 108; number of edges, 872; average node degree, 16.1; avg. local clustering coefficient, 0.697; enrichment p-value <1.0e-16.

Table (right side) - the table shows the nodes with the highest degree. In the table there are also reported CDH2, CTNNA1, and EPN1 just to show all seeds. The number of colored segments of each protein node shows in which of the dysregulated processes shown in **Table 8** it is involved.

 doi:10.20944/preprints202403.0473.v1

Table 8.

DYSREGULATED PROCESSES RELATED TO CYTOSKELETON					
Biological Process GO	Description	P	F.D.R.	Strength	Color
GO:2001237	Negative regulation of extrinsic apoptotic signaling path.	8.05	6.6e-06	1.18	
GO:0051129	Negative regulation of cellular component organization	7.23	3.79e-09	0.76	
GO:2000811	Negative regulation of anoikis	7.00	2.7e-04	1.58	
GO:0043069	Negative regulation of programmed cell death	6.66	3.24e-09	0.7	
GO:0060548	Negative regulation of cell death	6.50	4.93e-09	0.67	
GO:0048519	Negative regulation of biological process	5.62	2.67e-14	0.39	
GO:0043066	Negative regulation of apoptotic process	5.54	1.09e-08	0.69	
GO:0023057	Negative regulation of signaling	5.13	7.04e-08	0.58	
GO:0010648	Negative regulation of cell communication	5.11	6.42e-08	0.58	
GO:0031333	Neg. regulation of protein-containing complex assembly	4.56	6.4e-04	0.95	
GO:0010507	Negative regulation of autophagy	4.42	4.5e-04	1.1	
GO:0032369	Negative regulation of lipid transport	4.23	1.1e-03	1.39	
GO:2001234	Negative regulation of apoptotic signaling path.	3.74	2.5e-04	0.85	
GO:1902904	Negative regulation of supramolecular fiber organization	2.80	1.4e-03	0.89	
GO:0051494	Negative regulation of cytoskeleton organization	2.77	1.2e-03	0.9	
GO:0007162	Negative regulation of cell adhesion	2.31	1.2e-03	0.75	

The table shows the dysregulated processes in which these proteins are involved. Strength is in logs and for P (see Methods). The colors are reflected in the metabolic characteristics expressed by the various nodes in Figure 4.

We can conclude that interaction of SARS-CoV-2 ORF7b protein with host cell proteins, especially those involved in cytoskeletal modifications, plays a role in the virus's ability to propagate infect cell to target distant tissues. Structural disarrangements or metabolic dysregulations induced at the cytoskeleton level impact the cell's ability to counteract viral infection, aiding in viral spread, or facilitating intracellular transport of viral components, so contributing to its long-distance diffusion.

3.6. Topological Analysis

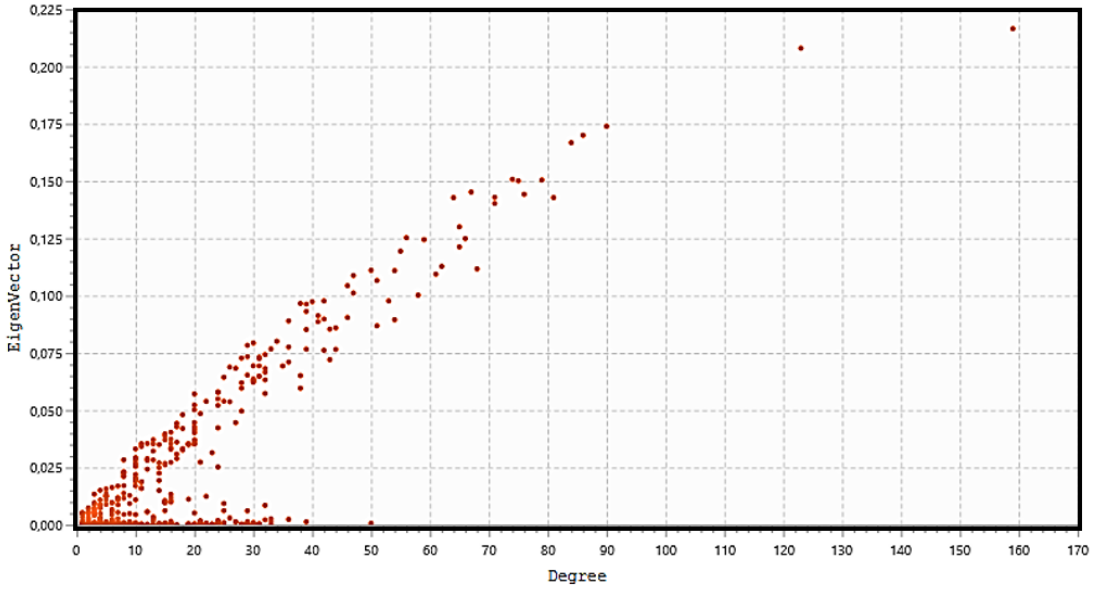
When a virus infects a cell, viral proteins represent the attackers and seek vulnerabilities in the network. Vulnerabilities introduce uncertainties into the network as a loss of original metabolic performance, even by changing information flows. Examining the network topology allows us to study both vulnerability and functional uncertainty, and to seek any architectural or functional changes. Crossing pathways between metabolic pathways or between signaling pathways are among the most vulnerable topologies, while hub-and-spoke topologies have the least uncertainty of destabilization. Therefore, topological data analysis is a powerful biological network analytic method [55]. To extract meaningful information from interactomic data, it is essential to understand the correlation between topological parameters and the mechanisms of biological functions [56]. Centrality metrics measure the importance of nodes by trying to quantify the idea that some nodes are more "important" than others.

We can roughly divide topology scoring metrics into two groups, the local one to evaluate individual nodes and the global one to evaluate the network. Global metrics include Betweenness, Bottleneck, Eccentricity, Closeness, Radiality, Stress and more. It is a useful methodological approach to increase the efficiency in selecting, characterizing and classifying crucial proteins as both hub and/or bottleneck proteins. In particular, bottlenecks are key link proteins, almost always not HUBs, but hard-to-discover essential proteins which control and regulate metabolic cross-overs. In fact, in regulatory networks, being intermediate (i.e., "bottleneck") is an indicator of functional essentiality, which is often much more significant than degree (i.e., of being a hub) for understanding the direction of an information flow.

Eigenvector Centrality measures the transitive influence of nodes. Relationships originating from high-scoring nodes contribute more to a node's score than connections from low-scoring nodes. If a node has a high eigenvector score, it means that it is connected to multiple nodes that have high scores as well. **Figure 5 (top)** shows the distribution analysis of the eigenvectors. The graph shows

 doi:10.20944/preprints202403.0473.v1

that the eight highest values have their degree value exactly matching that of the eight hub nodes previously selected, showing that all hub proteins also have the highest eigenvector scores. Stress is an index of node centrality. It represents the number of the shortest paths passing through a node. A high-stress node is a node traversed by a very large number of the shortest paths. In an interactomic network, it shows the relevance of a protein in keeping functionally communicating nodes together. We can consider such a protein as a "bottleneck" protein [57–59]. The higher its value in the network, the more relevant the protein is in linking regulatory proteins of different pathways. However, because of the parametric significance of this index, it is sometimes possible that stress shows only a molecule involved in many cellular processes but not relevant for maintaining communication between other proteins [60]. The F**igure 5** (**middle**) shows the stress distribution analysis where SEC13, EGFR, MTOR, HSPA5, VAMP2, and SRC are the major stress proteins. Betweenness [56] is also an index of node centrality, similar to stress, but with more information. It is a measure to rank the relative importance of vertices or edges. It represents the total number of non-redundant shortest paths connecting a pair of nodes, a1 and a2, crossing the node a. The betweenness value of a node increases if it lies on a non-redundant shortest path between nodes a1 and a2. Therefore, a high Betweenness score characterizes a key node in maintaining connections and this type of nodes becomes the critical point that controls the communication between other distant nodes in the network. In biological terms, it characterizes the interactivity of a protein in an interactome, showing the protein's ability to link distant proteins. Thus, betweenness is a measure of how important the node is to the flow of information through a network. This feature of the node in a protein signaling network may also show the relevance of the protein to act as a bottleneck. It acts as a junction connecting metabolic pathways that can hold the communicating proteins of different pathways together. The higher the value, the greater is the relevance of the protein as a bottleneck molecule. The interdependence of a protein effectively shows the ability of this protein to link distant proteins. In reporting modules, intermediate relationships are crucial to maintain the functionality and consistency of the reporting mechanisms.



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0473.v1

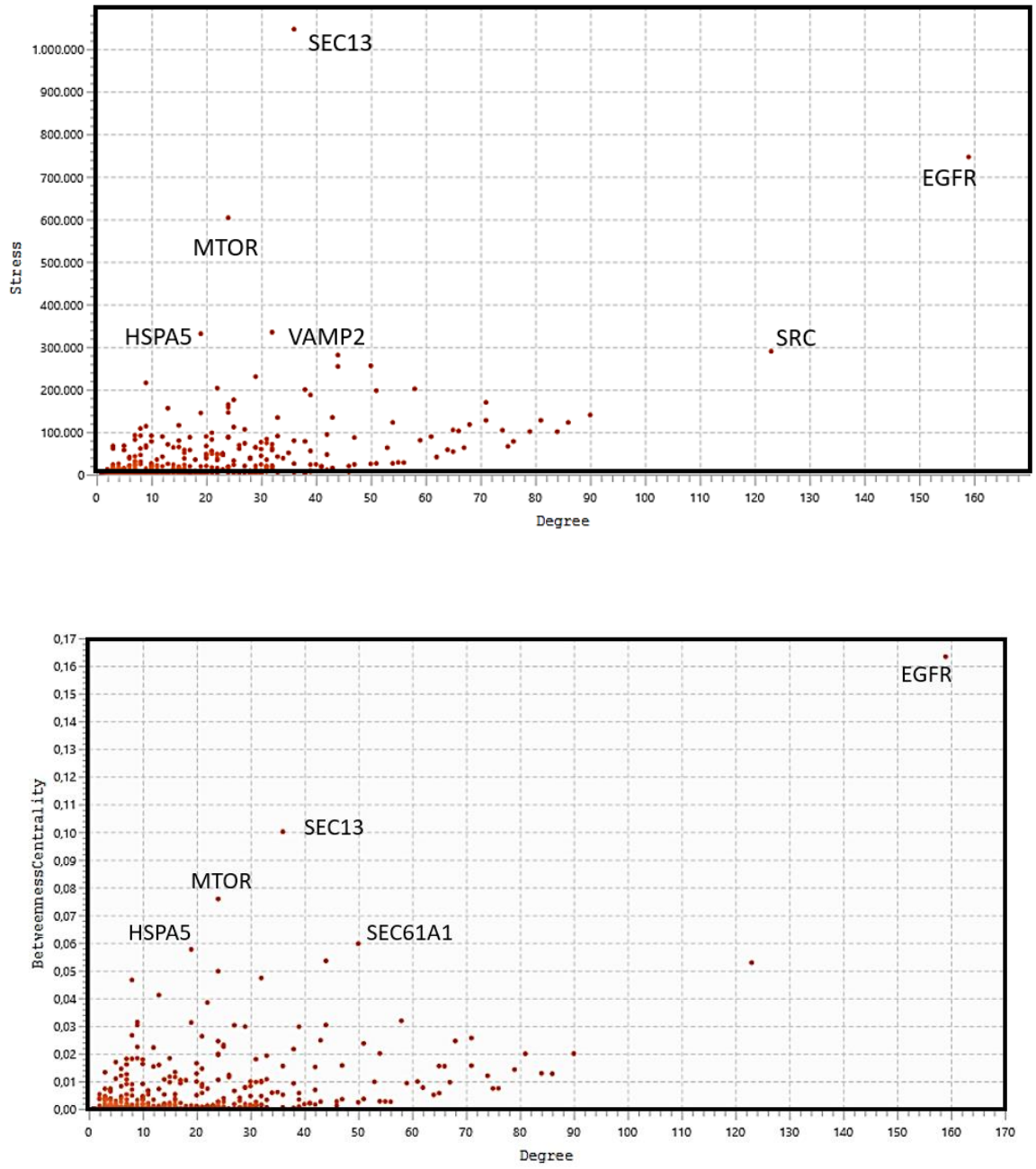


Figure 5. Eigenvector distribution (**top**); Stress distribution (**middle**); Betweenness Centrality distribution (**bottom**). We calculated distributions using Cytoscape with Analyzer and Centiscape. Cross-referencing parametric values have completed the selection of the best proteins in the Cytoscape Node Table for each protein.

The analysis in **Figure 5 (bottom)** confirms EGFR, SEC13, MTOR, HSPA5 as "bottleneck" proteins, also showing a new protein, SEC61A1. In the stress distribution, the SEC61A1 value was very close to that of VAMP2, while now is the VAMP2 value close to that of SEC61A1. Therefore, we can consider both proteins as bottlenecks.

Eigenvector, Stress, and Betweenness Centrality distributions were used in a multi-parametric approach to validate the 8 hub proteins and define the role of some proteins as bottlenecks. Among proteins selected as the most ranked bottlenecks (EGFR, HSPA5, MTOR, SEC13, SEC61A1, SRC, and

 doi:10.20944/preprints202403.0473.v1

VAMP2), EGFR and SRC show a dual role, both as a hub and as a bottleneck. Putting it all together, we have EGFR and SRC which are mixed (HUB/Bottleneck) proteins, HSPA5, MTOR, SEC13, SEC61A1, and VAMP2 which are pure bottleneck proteins, and PIK3R1, PIK3CA, GRB2, and HRAS which are pure hub proteins. These differences allow these proteins to be defined in three classes of molecular markers. In an eukaryotic protein interaction network, a node rarely represents the lone native protein because of alternative splicing [61] and proteoforms [62]. This may be a problem because in all databases (including STRING) it is customary to collapse all the dofferent functions of its isoforms and proteoforms onto the native protein, attributing to it a greater load of functions that it does not possess. In the interactome calculation, this anomaly produces biased nodes with higher and unreal connectivity.

Researchers have identified three different types of hubs in tissue-specific protein-protein interaction networks: few tissue-specific hubs, many tissue-preferred hubs that are formed by highly connected proteins, and housekeeping hubs that are involved in normal metabolic management [63]. When we connect these features to their specific functional roles within different tissues, they exhibit distinct functional differences that are influenced by the structure/function relationships.

Disordered regions significantly enrich pure hub and hub/bottleneck proteins among the three previous classes, and as a result, these proteins harbor a significant number of predicted binding sites [64]. They are also rich in splice variants, have longer peptide chains, and host a significant number of domains. This successful structural versatility drives their high propensity for interactions [62]. Because they are involved in essential functions such as phosphorylation and mRNA slicing processes, they get tangled in multiple intracellular functional pathways. Pure bottleneck proteins are typically extracellular proteins that are connected to pathological conditions, such as cancer, and play a role in cell-to-cell signaling pathways. Defining the actual functional role of a node is challenging because of the convergence of multiple functions with varying spatio-temporal characteristics. Many researchers still use static and deterministic approaches to select their experimental design, which leads to these limitations.

The topological role of network hubs depends also on the exponent value of the power law [65]. A value of <2 for the degree exponent b (see **Figure 3**), however, very close to 2, suggests a hub-and-spoke architectural model. The hub-hub network of the entire interactome fits a hub-and-spoke model, as Perera [66] and Barabasi [67,68] suggest. The largest hub (EGFR, 159 nodes) acts as a central coordinator and connects to a significant portion of nodes, which is shown in **Figures 3 and 5**. These structures act as a backbone connecting different metabolic modules. In this topological context, we should also identify the top-hubs as significant centers of control over the entire network. This view is also in agreement with the topological parameters calculated by the Cytoscape Network Analyzer.

The **Figure 6** also shows the relationships and the particular topology involving both HUBs and bottlenecks nodes [69]. **Figure 6S** (in the Supplements) shows how EGFR organizes in a topologically similar manner, even under normal conditions. Relationships between the HUB nodes are strong, while those with the bottleneck nodes are less intense, as the figure shows. All these significant nodes play a collective role in maintaining the stability of the hub-spoke system, albeit with varying functions and methods [70]. Each of them controls many and different biological processes [71]. The question remains: which node, regardless of its degree, is involved in the greatest number of functional processes? The question is not far-fetched. Because of the many metabolic crossroads, greater connectivity may not correspond to greater functional involvement [72]. When designing a drug, it is important to have this information.

 doi:10.20944/preprints202403.0473.v1

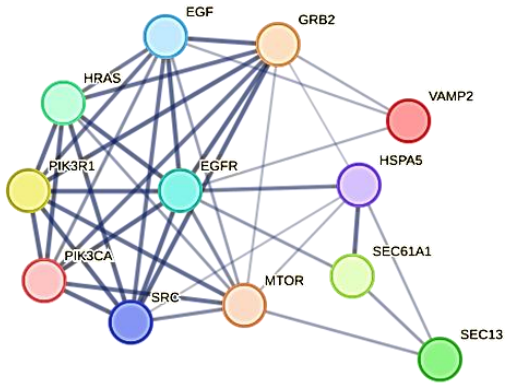


Figure 6. Hub-and-spoke organization of major HUBs in the ORF7b-induced human interactome. By removing all unnecessary nodes from the network in **Figure 2**, we extracted this graph. Edge intensity is proportional to the interaction intensity between nodes (calculated by STRING).

Table 9, while surprising for the very high number of functional involvements, shows how a HUB node is not always the main controller of the metabolic landscape. MTOR (degree = 24) and HSPA5 (degree = 19), although with lower connectivity, are involved in a very significant number of processes. The distribution of nodes and biological functions on the hub-and-spoke system, coupled with the ORF7b-induced interactome's complexity, handles this outcome. How functionally significant are the processes they regulate, would be the next inquiry. The answer would require a large analysis not covered by this study. Certainly, these same nodes, depending on their level of genomic expression, can both up-regulate and down-regulate a biological process [73–75]. Down-regulated processes, or "negative biological processes" according to GO, are important to highlight because of their higher probability of resulting from viral strategy [76]. Here, as we will see below, statistical significance is no longer the only parameter to follow.

Table 9. Involvement of HUBs and Bottlenecks in the control of Biological Processes (GO).

HUB protein	Number of GO Processes	Bottleneck protein	Number of GO Processes
EGFR	**408**	**EGFR**	**408**
PIK3R1	**328**	**HSPA5**	**234**
EGF	**646**	**MTOR**	**413**
HRAS	**245**	**SEC13**	**83**
GRB2	**233**	**SEC61A1**	**63**
SRC	**508**	**SRC**	**508**
PIK3CA	**271**	**VAMP2**	**143**

Note – EGFR and SRC are on both lists because of their dual activity. STRING extracted from the genes paired to each term the biological processes for each individual protein under the Biological Processes (GO) section.

3.7. The Functional Effects Depend Not Only on ORF7b but Also on the Integrated Action of Several Viral Proteins

The virus shows extraordinary strategic potential. Our previous results indirectly showed the specific impact of its proteins on crucial metabolic processes. About 200 symptoms of patients [77] generated various hypotheses based on clinical impression found to be associated with long-covid. All this shows how broad and diversified the systemic action of the virus is. Thus, part of the broad spectrum of metabolic activities found in this interactome might be associated with the multitude of clinically observed symptoms. However, we should not think that the ORF7b protein alone is capable of so much. The proteome yields biological functions via target proteins, which result from specific one-to-one interactions between viral and human proteins. Other viral proteins could target human

 doi:10.20944/preprints202403.0473.v1

proteins present in metabolic modules where ORF7b also operates. The ORF7b circular interactome (**Figure 1**) displays other viral proteins, ORF3a, and M, which may show their ability to target human proteins in the same metabolic modules as ORF7b. As of July 2023, we have organized a database called SHPID, which contains BioGRID interactions. In this database, we have collected 33,823 interactions between SARS-CoV-2 and human proteins. We analyzed the hub proteins highlighted in **Figure 3**. The proteins EGFR, SRC, and PIK3R1 are the major HUB nodes of the ORF7b interactome with 159, 123 and 90 links, respectively. Although these proteins are involved in the ORF7b interactome, **Table 10** reveals that they also interact with other viral proteins.

Table 10.

Viral Protein	Human target	Viral protein features**
nsp4*	EGFR	Is involved in the assembly of virally induced cytoplasmic double-membrane vesicles necessary for viral replication.
M*	EGFR	Component of the viral envelope.
ORF3a*	EGFR	Homotetrameric potassium sensitive ion channels (viroporin) and may modulate virus release
ORF7b*	EGFR	This paper
S	EGFR	Spike or Surface glycoprotein
nsp4*	SRC	See above
nsp5*	SRC	Is a cysteine protease, essential for the viral life cycle.
nsp6*	SRC	Plays a role in the initial induction of auto-phagosomes from host reticulum endoplasmic
nsp13*	SRC	Multi-functional helicase with a zinc-binding domain in N-terminus
nsp14*	SRC	3'-5' deoxyribonuclease
E*	SRC	Plays a central role in virus morphogenesis and assembly
M*	SRC	See above
ORF3a*	SRC	See above
ORF3b	SRC	Could be involved in immune evasion as interferon agonist (78)
ORF6*	SRC	Could be a determinant of virus virulence
ORF7a*	SRC	Non-structural protein, which is dispensable for virus replication in cell culture
ORF7b*	SRC	See above
ORF8	SRC	Is a viral cytokine regulating immune responses
S	SRC	See above
M*	PIK3R1	See above
ORF7b*	PIK3R1	See above
ORF3b	PIK3R2	See above
M*	PIK3R3	See above
S	PIK3R3	See above

 doi:10.20944/preprints202403.0473.v1

N*	ORF7b	Responsible for wrapping viral RNA into a symmetric helical structure

Notes *) With few exceptions (ORF3b, ORF8, and S) all the remaining viral proteins, although apparently compact, have intrinsically disordered regions (IDR), often in the tails, which make possible interactions with numerous partners. This could be the structural cause of their multiple action. **) Viral protein information from National Center for Biotechnology Information – National Library of Medicine – USA.

The **Table 10** depicts how these high-degree human proteins are a common target for many viral proteins. Our analysis of the interactions between the thirty-one viral proteins and the human proteome, as reported by BioGRID, yielded this result. Even though viral proteins have co-evolved with their human host or other species, they seldom possess structurally detailed molecular interfaces for accurate and stable interaction. Only a few viral proteins exhibit strong interactions, akin to those observed in complexes. Most of the interactions have weak bonds, also because of the anisotropy of the contact areas [79]. Viral proteins attempt to establish competition with normal binding proteins by mimicking interaction interfaces to the greatest extent possible, binding to target proteins with interaction constant values that typify weak processes. The interfaces mimicked by viral proteins compete through multiple and transient cellular interactions. They interact with hubs and bottlenecks in the human PPI network to control vital proteins in complexes and pathways. Proteins can overcome a structural difficulty by introducing an intrinsically disordered region (IDR) in the sequence, which can enhance the mimicry of contact surfaces. IDPs have IDR stretches that may be part of low affinity inter-molecular interactions [80]. With the emergence of IDPs in eukaryotic proteomes [81], the disorder becomes a crucial information for PPI evaluation.

Many of the interacting viral proteins in the **Table 10** show IDR (data not shown), thus, the probability of multi-targeting is high and this could explain the phenomenon (see Methods for details). After all, even the three human proteins analyzed have inherently disordered and highly mobile segments (data not shown). They are lipid-anchored proteins with the central body in the cytoplasm or outside the cell. Two long disordered and mobile tails are present in EGFR, which is found on several internal membranes (endosomes, ER, Golgi, nucleus) and on the surface. SRC also has long disordered and mobile tails and some mobile central segments and has multiple localizations, both on the surface and on intracellular organelles (endosomes, mitochondria, etc.). Finally, PIK3R1 too shows a long-disordered C term with many mobile intermediate segments and is on the cell surface. To this we should add that the disordered/mobile parts often show PTM sites. The presence of PTM sites expands the number of proteoforms for any single protein, increasing the probability of interacting with new molecular partners, establishing new functions.

A particular observation is that our database shows that ORF7b itself interacts with the viral N protein (see **Table 10**). Among the various functional peculiarities of this protein, we find it is involved in the formation of liquid droplets [178]. The liquid-liquid phase separation is considered the key mechanism for organizing macromolecules, such as proteins and nucleic acids, into membrane-free organelles [184], and N protein can self-bind into spherical aggregates which can freely diffuse in the condensed phase with liquid-like behavior [185,186].

Although we had also examined other relevant human HUB nodes of the ORF7b interactome, such as PIK3CA, EGF, and HRAS, we did not find other direct targeting of viral proteins. Therefore, these seem nodes extracted specifically from the ORF7b functional enrichment and functionally connected with the other HUBs of this network. Thus, their presence in this interactome seems due to a specific functional requirement of ORF7b. After all, the human metabolic system responds intricately to the ORF7b protein, consistent with the multiple metabolic responses of multicellular eukaryotic systems. In particular cases, viral action may require the synergistic action of different viral proteins. Thus, to achieve its biological effect, the virus can also use complex and sequential interaction modes on a single protein. This analysis is in excellent agreement with the previous classification of hub and bottleneck proteins. Unfortunately, we currently do not know where, how, and when these interactions occur. Hence, our vision of a dynamic phenomenon is only static and somewhat unclear, which may also be spatio-temporally inappropriate or distorted in our

 doi:10.20944/preprints202403.0473.v1

reconstruction of it [82]. Anyway, SARS-CoV-2 employs a known strategy of targeting the same human protein with multiple viral proteins [83].

3.8. The Peculiar Case of GRB2, a Protein in the Service of ORF7b

GRB2 (Growth Factor Receptor Bound Protein 2 – UniProt: P62993) is a protein that according to BioGRID binds ORF7b, although with the low level 1. While, our observation within the BioGRID dataset reveals that this protein exclusively interacts with ORF7b. We excluded it from the seed proteins owing to its low significance, but found it recluted in the interactome. The enrichment suggests that this ORF7b interactor is essential for virus infection. It assumes the role of HUB with 84 connections and controls 233 biological processes (see **Table 9**). GRB2 is an important protein that provides a critical link between the phosphorylated cell surface growth factor receptors (EGFR) and the PI3K-Akt signaling pathway. Both KEGG and Reactome Pathways reported its significant involvement in several signaling mechanisms (hsa04151, PI3K-Akt signaling pathway; HSA-1963640, GRB2 events in ERBB2 signaling; HSA-179812, GRB2 events in EGFR signaling; HSA-354194, GRB2:SOS linkage to MAPK signaling). Later on, we come to know that it often involves in various dysregulation processes that assist viral activity. **Table 9**'s proteins and GRB2's case show the sophisticated and diverse molecular strategy of SARS-CoV-2. The hubs listed in this table are proteins obtained through functional enrichment, but are not direct molecular interactors of ORF7b.

3.9. The Role of ORF7b

The diverse and sometimes contrasting metabolic properties of some of the interactome nodes are surprising. Among the 1,691 Biological Processes (GO) induced by ORF7b, there are 117 peculiar metabolic activities mentioned as negative activities (approximately 7%). Most of the HUB and bottlenecks proteins are also involved. According to AmiGO-2, the official web-based set of tools for searching and browsing the Gene Ontology database, negative activity means "any process that stops, prevents or reduces the frequency, rate or extent of metabolic functions". To identify which terms are most significant for these purposes, p-values alone cannot guide us. STRING measures the size of the enrichment effect using also the "Strength" score. The sole use of the p-value can produce an overrepresentation of the GO term, while the value of P (see methods) is useful for amplifying those underrepresented biological processes preferentially connected with a specific context [85] through their expression. A limitation of this approach is that, in a complex interactome, many proteins are not specific to a single metabolic pathway, but are sometimes even part of multiple pathways. Here, the massive study of some of these pathways favors the assignment of the protein to the more studied GO pathways. In fact, the databases favor assigning the protein to the more studied GO pathways and obscure the emerging relationships towards different biological pathways that are not studied or poorly represented [86]. Therefore, the analysis should select only the most reliable terms.

In addition, Hong et al., [86] demonstrated that functionally linked gene pairs, even in different functional pathway types, as defined in KEGG pathways, show positively correlated expression levels. Therefore, these two genes (or their proteins), even in a functional pathway altered by a disease, are similarly up-regulated or down-regulated. This is because of their reciprocal and close functional relationships [87]. So, when a disease affects a metabolic pathway, all the genes in the pathway will regulate their expression positively. Therefore, an over-representation of a GO process suggests an over-expression of the genes and their decoded products that make up the metabolic pathway, since they have close functional relationships with each other in regulating the expression [86,87].

We selected 17 terms with the highest possible strength value, paired with a very significant p value and listed according the value of P (see Methods). **Table 11** reports these terms according to the previously expressed rule. In the table, among the proteins involved in these negative functional activities, we can note (in bold) many of the proteins previously highlighted as HUB nodes, or as "bottlenecks" or involved in other important signaling pathways. Although all Biological Processes show positive values of enrichment (high strength), very many have minimal or negligible enrichments. It is necessary to exceed the value of 0.5 to have an enrichment of 3 times. We found

 doi:10.20944/preprints202403.0473.v1

that 32.28% of the processes have enrichments lower than 3 times and only 14.7% have enrichments greater than 10 times. The remaining 53.02% has intermediate enrichment values, between 3 and 10. This means that the most enriched fractions are very few and we can think the average enrichment of most biological processes as suitable for the normal metabolic function to be performed. The 17 selected terms therefore make up a very limited set, less than 1%, but the only one that can boast a statistically significant and even conspicuous enrichment. However, the negative term means over-enrichment and, therefore, suggests a gene over-expression. Some sets of proteins, enriching themselves, change their functional state, inducing changes in the pathways they control. Since the meaning of the negative term is loss of control, down-regulation, this means that by weakening their functions, they favor the activation, or deactivation, of the functional pathways they control. This is not new. We find a dysfunctional expression of genes with overexpression and deleterious functions during disease or even aging, in particular, of genes involved in pathways related to stress responses, antioxidant defenses, and DNA repair [88–90].

Table 11. Common altered pathways.

Function	Strength*	p-value	P	Human Proteins involved in the process
Negative regulation of ERBB signaling pathway	1.22	1.38e-18	22.13	HBEGF, EREG, PTPN12, TSG101, CBL, CBLB, **EGF**, ERBB2, CBLC, **EGFR**, TGFA, SOCS5, PTPN2, HGS, EPS15, ERRFI1, SNX5, SH3GL2, **GRB2**, BTC, AREG, SH3KBP1, CDC42, EPN1, EPGN
Negative regulation of EGFR signaling pathway	1.23	3.24e-17	21.53	HBEGF, EREG, TSG101, CBL, CBLB, **EGF**, CBLC, **EGFR**, TGFA, SOCS5, PTPN2, HGS, EPS15, ERRFI1, SNX5, SH3GL2, **GRB2**, BTC, AREG, SH3KBP1, CDC42, EPN1, EPGN
Negative regulation of anoikis	1.14	5.72e-05	6.56	**PIK3CA, ITGA5, BCL2L1, CAV1, PTK2, SRC, ITGB1**
Negative regulation of extrinsic apoptotic signaling pathway	0.76	9.26e-07	6.05	GCLC, LGALS3, BCL2L1, IGF1, CTNNA1, UNC5B, FYN, **FAS**, **CASP8**, LMNA, GCLM, **SRC**, AR, CTTN, NRG1, **ITGA6**, **AKT1**
Negative regulation of protein tyrosine kinase activity	0.99	9.13e-05	5.90	TSG101, CBL, CBLB, CBLC, SOCS5, PTPN2, **CAV1**, ERRFI1
Negative regulation of epidermal growth factor-activated receptor activity	1.18	1.7e-04	4.99	TSG101, CBL, CBLB, CBLC, SOCS5, ERRFI1
Negative regulation of interleukin-6 production	0.81	5.0e-05	4.61	CSK, SOCS5, GAS6, TLR9, VIMP, PTPN6, ARRB1, ENSP00000417517
Negative regulation of peptidyl-tyrosine phosphorylation	0.88	1.50e-05	4.55	TSG101, CBL, CBLB, CBLC, SPINK1, SOCS5, PTPN2, CAV1, ERRFI1, PRKCD, PTPN6
Negative regulation of PERK-mediated unfolded protein response	1.33	9.2e-03	4.12	NCK2, PTPN1, NCK1
Negative regulation of endoplasmic reticulum unfolded protein response	1.04	8.9e-03	4.11	NCK2, **HSPA5**, PTPN1, NCK1
Negative regulation of blood-brain barrier permeability	1.55	3.13e-02	3.88	SH3GL2, VEGFA
Negative regulation of response to oxidative stress	0.81	4.1e-04	3.74	SLC7A11, MET, GGT7, CTNNB1, FYN, NFE2L2, **INS**, HIF1A, **AKT1**
Negative regulation of protein tyrosine phosphatase activity	1.39	4.67e-02	3.71	LGALS3, GNAI2
Negative regulation of mesenchymal to epithelial transition	1.38	4.77e-02	3.69	CTNNB1, STAT1
Negative regulation of blood coagulation	0.83	2.8e-04	3.69	PROC, PDGFRA, F2, PLAUR, PLAU, EDN1, CD9, PROS1, PRKCD

 doi:10.20944/preprints202403.0473.v1

Negative regulation of primary miRNA processing	1.38	4.67e-02	3.68	STAT3, IL6
Negative regulation of lipid transport	0.79	1.74e-02	1.77	**EGF**, PTPN11, SREBF2, AKT1, ITGB3

Note: *) For comparisons, a strength of 1.39 = 24.5 times enrichment, and 0.76 = 5.8 times. In red **bold**, the proteins highlighted in the text as HUB nodes, or as "bottlenecks" or involved in other important signaling pathways.

Our examination of **Table 10** enables us to confidently affirm that many pathways show statistically significant dysregulation, and we may have successfully identified pivotal genes associated with these pathways. At present, accurately describing what occurs is challenging because of the lack of data to pinpoint causes, determine the opportune moment for the process, and establish the sequence of events, all because of the absence of space-time information. The strategy of ORF7b, in collaboration with other viral proteins, aims to create a viral microenvironment that helps infected cells minimize cell matrix rigidity and adhesion, increase intracellular oxidative stress, generate pro-survival signals, to trigger the epithelial-mesenchymal transition process, to inhibit intracellular transport and ER activity, starting widespread cellular metabolic deregulation. We should emphasize that the process of metastasis, characterized by the epithelial-mesenchymal transition (EMT) and its inverse, the mesenchymal-epithelial transition (MET), plays a crucial role in the metastatic spread of carcinomas [91]. Likewise, these events appear to be among the primary targets in preventing programmed cell death mechanisms of infected cells, allowing survival after separation and systemic spread.

In particular, we can see the dysregulation of all protein tyrosine kinase receptor activities. This reduces the processes of internalization of external signals and the activities of receptors activated by growth factors. Integrin-mediated alterations of the intercellular matrix and loss of control over cell-extracellular matrix adhesion processes are also favored by integrated dysregulation of oxidative stress, unfolded protein response of the ER and lysosomal action [92,93]. The intention behind all these activities is to dysregulate programmed death processes such as apoptosis and anoikis, promoting the spread of infected cells in the body.

The systemic spread of infected cells explains well why the tissues and organs showed as infectible in **Table 6** are so numerous and all significant. In the presence of infected cellular material widespread in the body, the virus has also the potential ability to cause inflammatory processes in the brain, so it is important to pay particular attention to the dysregulation of blood-brain barrier permeability. Through altering endocytosis, endosomal trafficking, lysosomal degradation, blocking anabolic processes and lipid transport, this creates mitochondrial dysfunction, resulting in a heavy dependence on glucose for energy production. Numerous miRNAs work within the cell and could interfere with these procedures. However, distinguishing them individually through this type of analysis is not yet possible.

In a nutshell, this tiny protein is involved in controlling the intercellular communication of the virus. By suppressing intracellular signaling, it created a metabolic microenvironment that caused generalized metabolic dysregulation and blocked intracellular transport of cargos. Prevention of local programmed death mechanisms leads to viral shedding. Various viruses show comparable infection strategies [95], such as extending particular stages of the cell cycle, managing programmed cell death, and using the nuclear membrane to transmit viral genetic material to and from the nucleus. These findings help to understand how SARS-CoV-2 can spread via cell-to-cell transmission [95], where ACE2 is not required. Our assessment shows that viral mutations shared by different variants are unsuitable for evaluating disease mechanisms. This is due to the high metabolic interference capacity of the remaining information package of the virus. Attention to mutations in the Spike protein has distracted from the evaluation of the molecular mechanisms underlying the metabolic dysregulations induced by the virus.

3.10. Cluster Analysis

 doi:10.20944/preprints202403.0473.v1

Cluster analysis allows us to extract protein interaction sub-networks that interact with each other in functional complexes and pathways to produce reliable hypotheses that can explain the various dysregulations of human metabolism induced by ORF7b. This also increases the likelihood of identifying candidate genes/proteins that can help us understand the rationale for viral action and the metabolic pathways involved.

Cluster analysis is a data analysis that explores the groups present within a dataset, known as clusters. We used Cluster K-means analysis, which does not need to group data points into predefined groups and is an unsupervised learning [96] method. In unsupervised learning, insights come from data without predefined labels or classes. K-means is also an iterative partition algorithm and is a good clustering algorithm that ensures high similarity within cluster and low similarity between clusters. The clusters representing our entire population of interacting molecules in the ORF7b interactome derive from a base of significant experimental data and rigorous procedures for implementing the network. This should produce high-quality clusters, which means non-redundant and low-noise results, as they can reduce the quality and interpretability of the clusters. The value to be attributed to K is one of the major drawbacks of this algorithm. In our analysis, K is equals 10 (**Figure 7**).

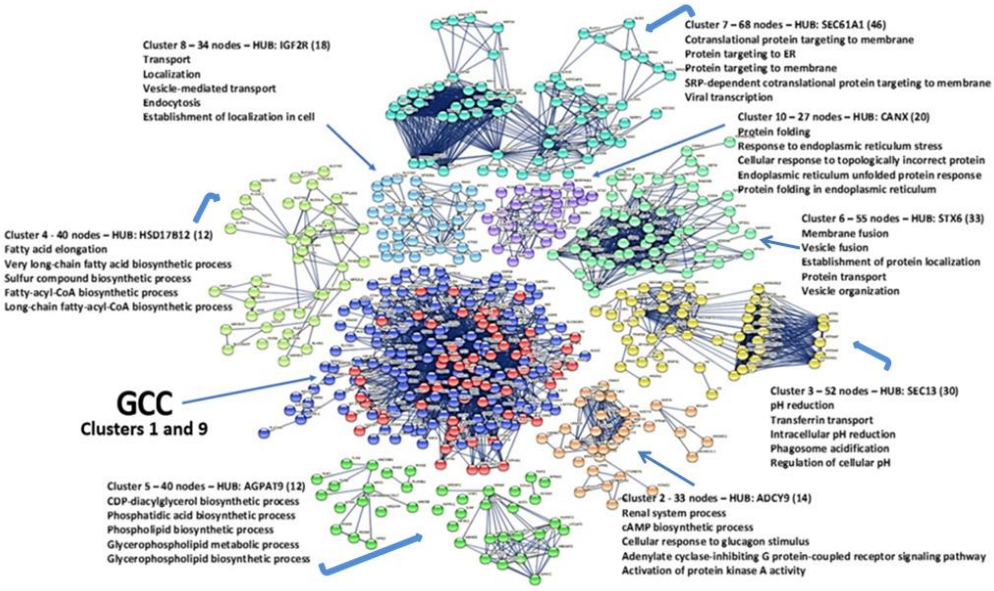


Figure 7. Clustering. The analysis shows ten clusters, all clearly identifiable except for the two central ones. All 10 clusters are statistically significant with p-values < 1.0e-16. In brackets, next to each key hub, there is its degree. We have not highlighted the links between the clusters to make them clearly visible. The Giant Connected Component (GCC) is made by two overlapping central clusters which add up to 206 total nodes, approximately 37% of the entire interactome. Except for clusters 1 and 9, which have distinctive features and require separate treatment, the most crucial parametric information is next to each cluster.

This result, got after many attempts with lower K values, has to be considered as the best compromise. We used this K-value because it gave us the most compact clusters and statistically significant p-values (all p-values are always <1.0e-16). The ten metabolic modules are all functionally consistent, and in the **figure 7S** and **8S**, we also show the links existing between the clusters. The many metabolic relationships existing between the clusters, as shown in the figure, mostly represent the normal metabolic machineries necessary for cellular life. Only the GCC shows an overlay of two modules, but, as we shall see, they resolve into two independent sub-graphs. The greatest interest is precisely in these two sub-graphs because they contain most of the HUBs and bottlenecks nodes previously found and control crucial metabolic pathways. While the other sub-graphs seem to

 doi:10.20944/preprints202403.0473.v1

regulate typical metabolic activities, understanding the specific functions of these central modules and where their constituent proteins operate within the cell is essential. This is a core-periphery organization. Core-periphery is a characteristic we can find at group-level relationships in biological networks, but not only [97]. The situation involves meso-scale dominance events [98]. It describes a scenario where a group of core nodes captures an excessive number of contacts in the network. On the contrary, the nodes on the periphery possess fewer interconnections with one another, albeit they are connected to the core nodes. In networking, the mesoscale describes sub-cellular events on length scales ranging from that of a single cell, up to the size of molecular complexes, where groups of molecules self-organize relationally to form large, functional core structures [99]. While individual nodes perform only local operations, their organization into clusters generates a richer and more diverse functional repertoire.

3.11. Analysis of GCC Core

The cluster analysis extracted from the compact GCC area two clusters (1 and 9), both statistically significant and compact. **Figure 8** shows the cluster No.1. In the caption, there are the major topological parameters. This cluster is very compact. Its major role is to regulate the EGFR family signaling pathway (EGFR, ERBB, ERBB2) where the receptors' protein tyrosine kinases signaling show a p <6.85e-48. It is involved in regulation of the Jak-Stat pathway, ERBB and ERBB2 signaling (p <2.55e-40), and regulation of peptidyl-tyrosine (p <2.99e-27). We can find the key details in the following GO terms: GO:0007169, GO:0038127, and GO:1901184. But in the cluster No1 we find also ITGB1, CAV1, EGF, EGFR, PIK3CA, INS, GRB2, PRKCA, HRAS, MTOR, just to mention the major nodes. Thus, the role of this cluster is also to control cell migration, cell motility, immune response, phosphorylation, cell death, apoptotic cell process, cell adhesion, cell migration, stress, insulin path, phagocytosis, lymphocyte activation, blood coagulation, Cytokine-mediated signaling pathway with very high statistical significance, as it appears from the list calculated by STRING in the Biological Process (GO) category.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0473.v1

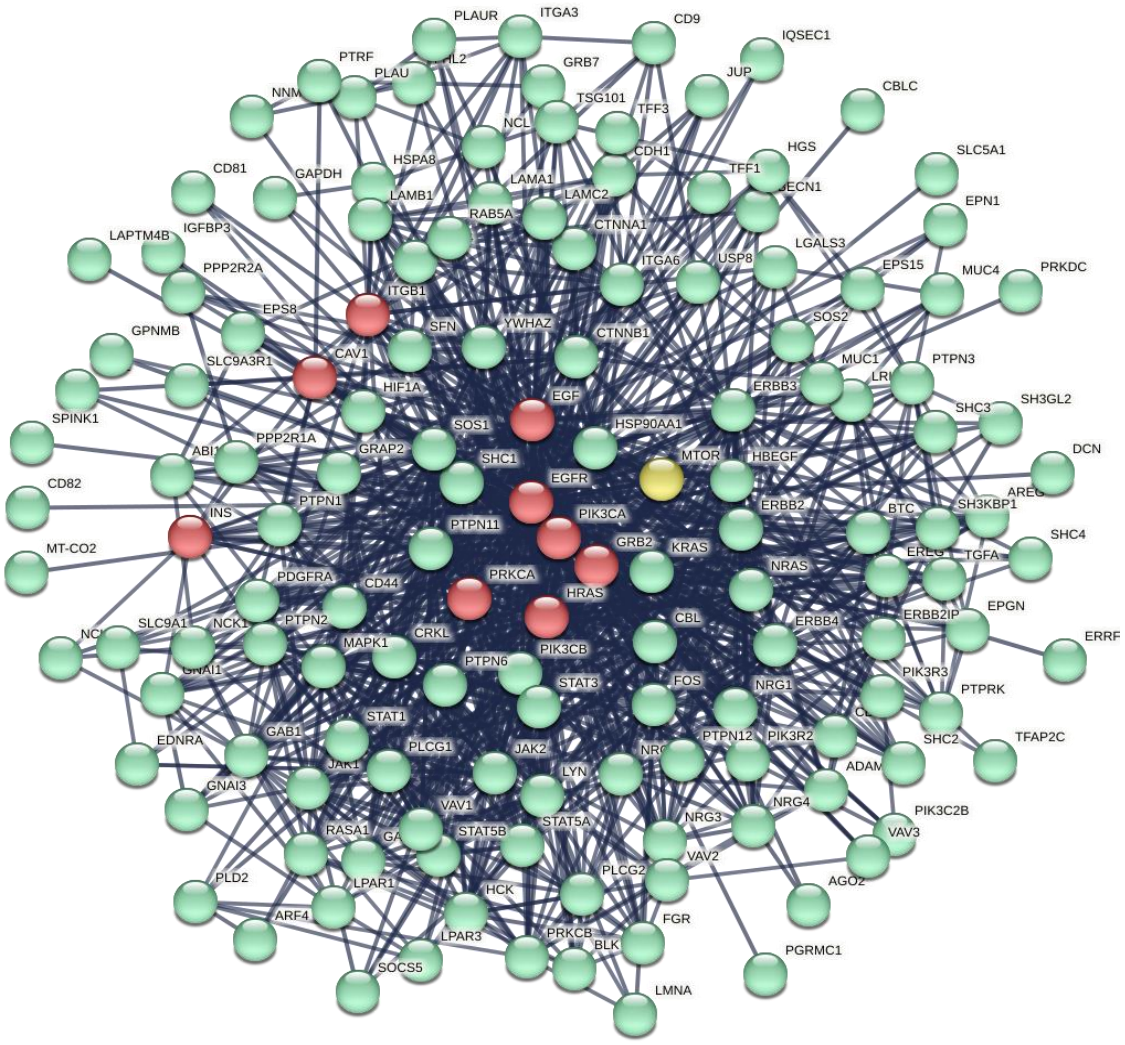


Figure 8. Cluster No.1 – 140 nodes, 1110 edges, p-value <1.0e-16. Average node degree 15.4, avg. Local clustering coefficient 0.622 (expected No of edges in a similar random network, 202), network diameter 3, network radius 2, Characteristic path length 1.91, network density 0.108. Main HUB node, EGFR (degree = 123). In red, HUBs previously found in the whole net; in yellow, a bottleneck node.

Proteins operate in their specific environments, therefore knowledge of where proteins are located is crucial to understanding the metabolic processes of which they are a part. We can perform this analysis with the help of Cytoscape. After transferring the cluster 1 to Cytoscape, we have with the help of STRING app and Nodes Table (Compartment analysis) selected the protein nodes with the highest statistical value (5.0) that operate in the various cellular compartments. Level 5 collects the most important proteins in defining the biological processes of which they are part.

In the **Table 11** we can see in which cellular compartment the cluster No 1 proteins operate, but we also see that there are various proteins already defined as dysregulated, so we can know where they operate. Nucleus and plasma membrane, as well as the cytoskeleton, are among the richest compartments of functional activities and proteins crucial for the progression of these activities. In the T**able 11** we find many of these proteins, for which symbolic notations have been used to distinguish them (see note to the **Table**). The table summarizes two important proteomic characteristics: a) there are numerous proteins that operate in a multipolar way, i.e., in several compartments (e.g., EGFR); b) there are many dysregulated proteins, in particular those involved in the fundamental processes of signaling and in favoring cell diffusion. Various proteins localize in

 doi:10.20944/preprints202403.0473.v1

multiple compartments, showing a shared protein pool even if apparently unrelated. However, each protein has its own level of expression and its own compartmental distribution.

Viewing all together, we could read this as indicative of functional progressions starting at the membrane and proceeding towards the nucleus. The limit is the absence of temporal information that statistically flattens the metabolic dynamics and makes it very difficult to have reliable sequential explanations. But this is not the only intricacy. **Figure 7S and 8S** demonstrate how single nodes can take on multiple roles to engage in various functional processes. Even a single functional activity can have its nodes distributed in numerous modules. This is a straightforward demonstration of how difficult it is to describe the actual behavior of concurrent functional processes without a temporal chronology, but the entire network, i.e., the operational context, can help.

Not only the regulation of space and time but also the compartmentalization characterizes the cellular proteomes. The presence of similar proteins in different compartments suggests different local proteomes [100], each performing its local metabolic activities, so it is difficult to identify any distortion. Nucleus and cytoplasm are among the most populated compartments. The proteomes of these compartments show a multipolar protein distribution, which makes them functionally very ductile. Therefore, attributing static and specific roles to the metabolism and to the proteins that operate within is a vision that does not correspond to reality. We cannot attribute a protein's metabolic function solely to its presence or absence. The function is also determined by the reactions happening at different omic levels and compartments [101]. Reactions that are always the result of protein-protein interactions. Thus, interactomic level reflects what happens at the genomic or transcriptomic level, generating a network that differs from the underlying ones because displays a portion of the total functional mechanisms. The event in question has recently gained prominence [102]. Some melanoma cells show a dependence on external sources of methionine for their growth. The authors describe the methylome, transcriptome and proteome of these cells. Only the multilevel contemporary study allowed the authors to understand the real metabolic behavior of methionine addiction because the study of the methylome alone led to trivial conclusions.

In short, we have the spatial distribution of proteins of the ORF7b interactome, but the temporal distribution is missing. Multi-localization of a protein increases the probability of interactions, generating possible new functional characteristics specific to the context. This expands the functional capabilities of the cell but makes any modeling that does not include all the parameters involved difficult.

Due to functionally important proteins, cluster No 9 has the potential to perform multiple functions (**Figure 9**). This cluster controls the process that modulates the cell transport to, or maintained in, a specific location (GO:0032879 p = 2.30e-34); extent of addition of phosphate groups to a molecule (GO:0042327, p = 1.89e-29); cell migration (GO:0030334, p = 3.11e-29); regulation of cell migration (GO:0030334, p = 3.11e-29); The transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169, p = 1.78e-28). It is also associated with the negative regulation of cell death (GO:0060548, Strength = 0.92, p = 1.75e-17) and programmed cell death (GO:0043069, Str.= 0.90, p = 3.96e-16). 0.98, p = 2.26e-16) or in the Negative regulation of production of miRNAs involved in gene silencing (GO:1903799, Str. 1.78, p = 4.6e-4). Similar considerations also apply to cluster No 9.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0473.v1

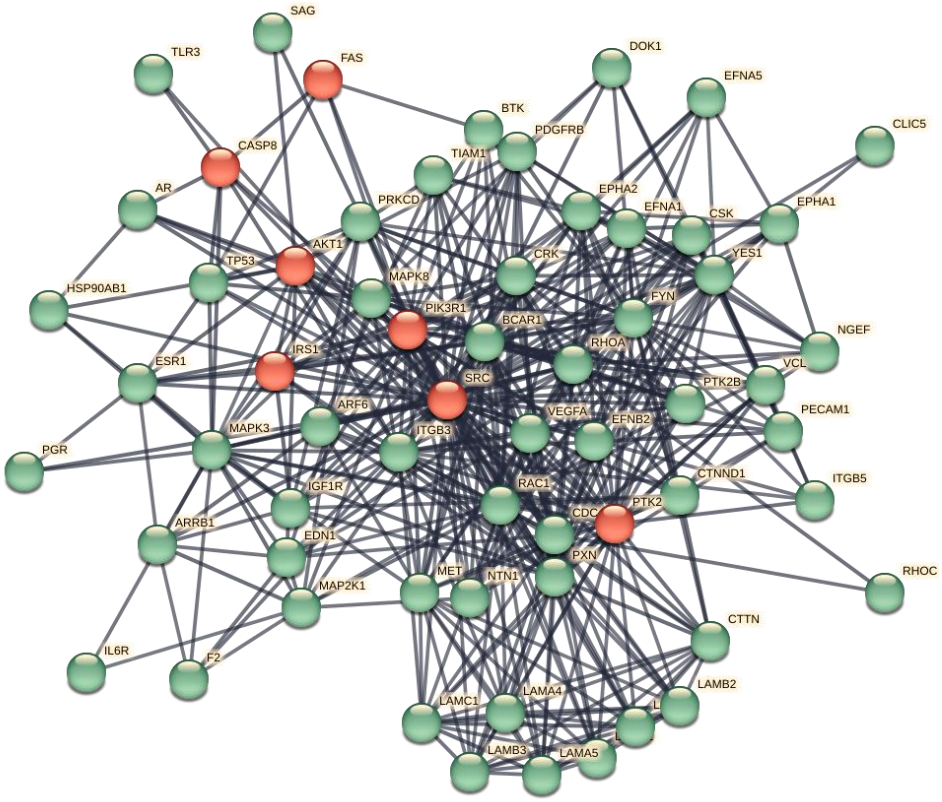


Figure 9. Cluster No.9 – 62 nodes, 437 edges, p-value <1.0e-16. Average node degree 14.097, avg. Local clustering coefficient 0.682 (expected No of edges in a similar random network, 83), network diameter 3, network radius 2, Characteristic path length 1.798, network density 0.231. Main HUB node, SRC (degree = 56). In red, some of the principal nodes of this cluster.

Dysregulated proteins, such as CTNNB1, SRC, PTK2, ITCB3, or PRKCD, found in Cluster No 9 (see **Table 12**), are present in many cellular compartments, including those that are distant from each other or different from a chemical-physical point of view, such as cytosol and plasma membranes. This means that they regulate temporally their expression and that they require post-translational modifications that depend on the context. Because most analysis platforms collapse this information into the native protein, nodes end up having more functional links than the context. This induces errors on the degree value and on the related topological evaluations, which can lead to alterations in the network.

An instance of this is the activation of the Human SRC (P12931, Proto-oncogene tyrosine-protein kinase Src), a Non-receptor protein tyrosine kinase that is triggered upon binding to various cellular receptors, including integrins and other adhesion receptors, regulating a wide range of biological processes. It belongs to the Src kinase family and is functionally redundant, making it challenging to identify its specific role in each compartment and determine which member is involved without the knowledge of its spatio-temporal characteristics in that specific context.

 doi:10.20944/preprints202403.0473.v1

Table 12. Operational cellular compartments of cluster No1 proteins.

COMPARTMENT	PROTEINS*	Protein number
EXTRACELLULAR	AREG, **BTC**, **CD81**, CD9, EGF, **EGFR**, **ERBB3**, **EREG**, **HBEGF**, **HSPA8**, INS, LAMA1, LAMB1, **MUC1**, ***NRG1***, **NRG3**, PLAU, SFN, **TGFA**, **TSG101**	20
CYTOSKELETON	***CTNNA1***, ***CTNNB1***, **GNAI1**, **GNAI3**, ***LMNA***, **MAPK1**, **PPP2R1A**, **PTPN3**	8
PLASMA MEMBRANE	ADAM17, **ARF4**, **BTC**, ***CAV1***, **CAV2**, CD44, **CD81**, CD82, **CDH1**, ***CTNNA1***, ***CTNNB1***, EDNRA, **EGFR**, EPS15, **ERBB2**, **ERBB2IP**, **ERBB3**, **ERBB4**, **EREG**, GAB2, **GNAI1**, **GNAI3**, **HBEGF**, **HCK**, **HRAS**, ITGA3, *ITGB1*, ITGB4, JUP, KRAS, **LAPTM4B**, **LPAR1**, LPAR3, **LYN**, **MUC1**, ***NRG1***, **NRG3**, PDGFRA, **PIK3C2B**, **PLCG1**, PLCG2, **PPP2R1A**, **PRKCA**, **PRKCB**, **PTPN2**, **PTPN3**, PTPRK, **PTRF**, SHC1, SLC9A1, SLC9A3R1, **TGFA**, **TSG101**, **USP8**	54
CYTOSOL	**PIK3C2B**, GRB7, **ARF4**, **PPP2R1A**, **PLCG1**, **HCK**, **USP8**, **PRKCA**, **MAPK1**, **RAB5A**, **FOS**, **HSPA8**, ***CTNNB1***, **HIF1A**, GAPDH	15
MITOCHONDRION	**PPP2R1A**, **MAPK1**, HSP90AA1, ***LGALS3***, **ERBB4**, **PTRF**, MT-CO2	7
GOLGI	**CAV2**, CBL, **CDH1**, **HRAS**, **LYN**, **MAPK1**	5
ER	**FOS**, **NCK1**, **PTPN2**	3
PEROXISOME	No level 5 protein	-
ENDOSOME	**CDH1**, **EGFR**, ***CAV1***, **ERBB2**, PTPN1, **RAB5A**, **MAPK1**, **TSG101**, **GRB2**, HGS, **USP8**, **LPAR1**, **LAPTM4B**, **GRAP2**	14
LYSOSOME	**LAPTM4B**, **HSPA8**, MTOR, **HCK**	4
NUCLEUS	**CAV2**, ***CTNNB1***, **EGFR**, **ERBB2**, **ERBB2IP**, **ERBB4**, **FOS**, **GRAP2**, **GRB2**, **HIF1A**, **HRAS**, **HSPA8**, IGFBP3, JAK2, ***LGALS3***, ***LMNA***, **LYN**, **MAPK1**, **MUC1**, **NCK1**, NCL, ***NRG1***, **PLCG1**, **PPP2R1A**, **PRKCB**, PRKDC, PTPN11, **PTPN2**, PTPN6, **PTRF**, *STAT1*, STAT3, STAT5B, TFAP2C	34

Note: *) Only proteins with the highest statistical significance value of 5 according to Cytoscape (values range from 0 to 5). We calculated protein node compartmentalization and values in Cytoscape using the STRING app. Highlighting all the proteins in Table IX would have rendered this table unreadable. 1) **In bold black**, the proteins which are present in more than one compartment. 2) We have identified and **underlined** the proteins responsible for the dysregulation of ERBB signaling, EGFR, protein tyrosine kinase activity, and regulation of peptidyl-tyrosine phosphorylation, as shown in TABLE IX. 3) Proteins also involved in deregulating apoptosis and anoikis to allow diffusion are also **in red italics.** Proteins in bold black and underlined or red italics and bold are common to two groups.

Table 13. Operational cellular compartments of cluster No9 proteins.

COMPARTMENT	PROTEINS*	Protein number
EXTRACELLULAR	EDN1, **F2**, **FAS**, HSP90AB1, LAMA5, LAMC1, **MET**, NTN1, **VEGFA**	9
CYTOSKELETON	CDC42, **CTNNB1**, CTTN, LMNA, MAPK3, **PTK2**, PXN, YES1	8
PLASMA MEMBRANE	**AKT1**, ARF6, , **CASP8**, **CAV1**, **CDC42**, CDH1, CDH2, **CTNNB1**, CTNND1, EFNA5, EFNB2, EPHA1, EPHA2, ESR1, **FAS**, HRAS, IGF1R, **ITGB3**, **MET**, NEDD4, PDGFRB, PECAM1, **PRKCD**, **PTK2**, PTK2B, PXN, RAC1, RHOA, **SRC**, TIAM1, TJP1, YES1	32
CYTOSOL	**AKT1**, ARF6, **CASP8**, **CTNNB1**, MAPK3, **PRKCD**, **PTK2**, RHOA, **SRC**, YES1,	10
MITOCHONDRION	GJA1, HSP90AA1, MAP2K1, MAPK3, **SRC**	5
GOLGI	**CBL**, CDH1, ESR1, HRAS, MAP2K1, MAPK3, NEDD4, RAC1, YES1	9
ER	**PRKCD**, MAP2K1	2
PEROXISOME	No level 5 protein	- -
ENDOSOME	ARF6, **CAV1**, CDH1, MAP2K1, MAPK3, **PRKCD**, RAC1, **SRC**	8
LYSOSOME	PDGFRB, PRKCD, **SRC**	3

 doi:10.20944/preprints202403.0473.v1

NUCLEUS	**AKT1**, **AR**, **ARRB1**, **CTNNB1**, ESR1, GJA1, HRAS, HSP90AB1, **ITGB3**, **LMNA**, MAP2K1, MAPK8, NEDD4, PGR, **PRKCD**, **PTK2**, PTK2B, RAC1, **STAT3**	19

Note: *) Only proteins with a statistical significance value of 5, according to Cytoscape analysis. These values range from 0 to 5. We calculated protein node compartmentalization and values in Cytoscape using the STRING app. In **bold**, the proteins in common with **Table 9**.

3.12. Co-Regulation between Hub and Bottleneck Proteins, Transcription Factors and miRNAs

Our findings thus far, have revealed a metabolic depiction that outlines the involvement of a specific group of significant high-ranking proteins in a series of dysregulated metabolic processes aimed at promoting the dissemination and spread of virus-infected cells throughout the body, because of the influence of the accessory viral protein, ORF7b. However, we still have limited vision because we can only glimpse the purposes, know some of the involved actors, but we still cannot understand who planned and performed the entire process.

Understanding the intracellular mechanism of complex biological processes driven by ORF7b also depends on deciphering its complicated co-regulatory network. The identification of Hub and Bottleneck proteins in protein groups dysregulated by viral infection prompts investigation to understand their co-regulation. Within the co-regulatory network, there are both post-transcriptional and transcriptional regulators that can regulate themselves and each other.

A limitation that should give pause for thought is the evidence that hub and bottleneck proteins control and regulate an enormous number of functional processes. Finding them involved in a particular process, even experimentally, does not mean that process is actual and existing in fact. Precise rules govern the occurrence of a functional process, primarily depending on the context of the events and the chemical-physical characteristics of the compartmentalized microenvironment where the event should occur. To ensure a functional event, the cell must program when, where, and how it should occur. The metabolic network is not solely dependent on proteins. To synchronize basic functional activities according to the circadian cycle or unexpected events, we need several other actors to accelerate or slow down an intricate and dynamic system. The comprehension of co-regulatory mechanisms that are fundamental to cellular identity and function requires the involvement of transcription factors (TFs) and microRNAs (miRNAs). TFs and miRNAs work together to regulate transcription and post-transcriptional processes, respectively [102,103].

Combining computational and experimental interaction data in network models can highlight functional mechanisms in TF- and miRNA-mediated gene regulation. These models can provide insight into the mechanisms that control gene expression at the system level, rather than at the individual gene level. Typically, TFs act as activators or repressors, increasing or decreasing transcription, while miRNAs are mostly repressors. We can visualize the distinct activities by using two separate networks: transcriptional networks and post-transcriptional networks. It is noteworthy that both networks are bipartite and direct. In each network, there are two distinct types of nodes interconnected by unidirectional edges. One network contains interactions between genes and transcription factors, which is known as a transcriptional regulatory network. The other network contains interactions between genes and miRNAs, which is known as a post-transcriptional regulatory network. We assume that, in post-transcriptional regulations, the regulatory actions of miRNAs towards targets are negative. However, it is possible to get integrated gene regulatory networks that include genes, TFs, and miRNAs, provided that the components are statistically more significant. The databases on TFs and miRNAs are quite recent and the data collected are both experimental and predictive because this area of research is still very young. Selective filtering is required to get statistically significant nodes. As explained in the Methods section, the reference databases of transcriptional and post-transcriptional networks comprise experimental data, whereas the integrated co-regulatory database comprises mixed data. This means that the comparison of the integrated co-regulatory network with the transcriptional networks may yield diverse interactions, which depend on the respective node rank in the two distinct systems.

 doi:10.20944/preprints202403.0473.v1

3.13. Transcriptional and Post-Transcriptional Regulatory Networks

As a result, in transcriptional regulatory networks, FTs possess two types of action since it is the TF that binds to its target gene, rather than the reverse. The information comprises an in-degree, which signifies the number of transcription factors binding a gene, and an out-degree, which signifies the number of genes bound by a transcription factor. All this reflects the functional and biological aspects underlying these interactions. High-grade TFs (i.e., hub TFs interacting on many genes) have a high key character of biological functionality, while target genes bound by many TFs do not have a tendency to be functionally essential. Therefore, analyzing this type of network provides insights into biological systems that are not obtainable through single gene studies.

Both the networks containing TFs and miRNAs are represented in **Figure 10**, illustrating the transcriptional and post-transcriptional networks of gene interactions, which include hubs and bottlenecks. The transcriptional network reveals that EGFR, the top-ranking hub node within the PPI network, possesses an in-degree value of 1 in relation to its interaction with ZNF263, whereas ZNF263 exhibits an out-degree value of 3. Therefore, within this network, ZNF263 holds greater biological significance in relation to EGFR. Its role in this transcriptional network involves functioning as a DNA-binding transcriptional repressor that specifically targets RNA polymerase II, resulting in the repression of EGFR, PIK3R1, and VAMP2. The TFs and miRNAs represented in the two networks are those of higher rank with a higher probability of interaction.

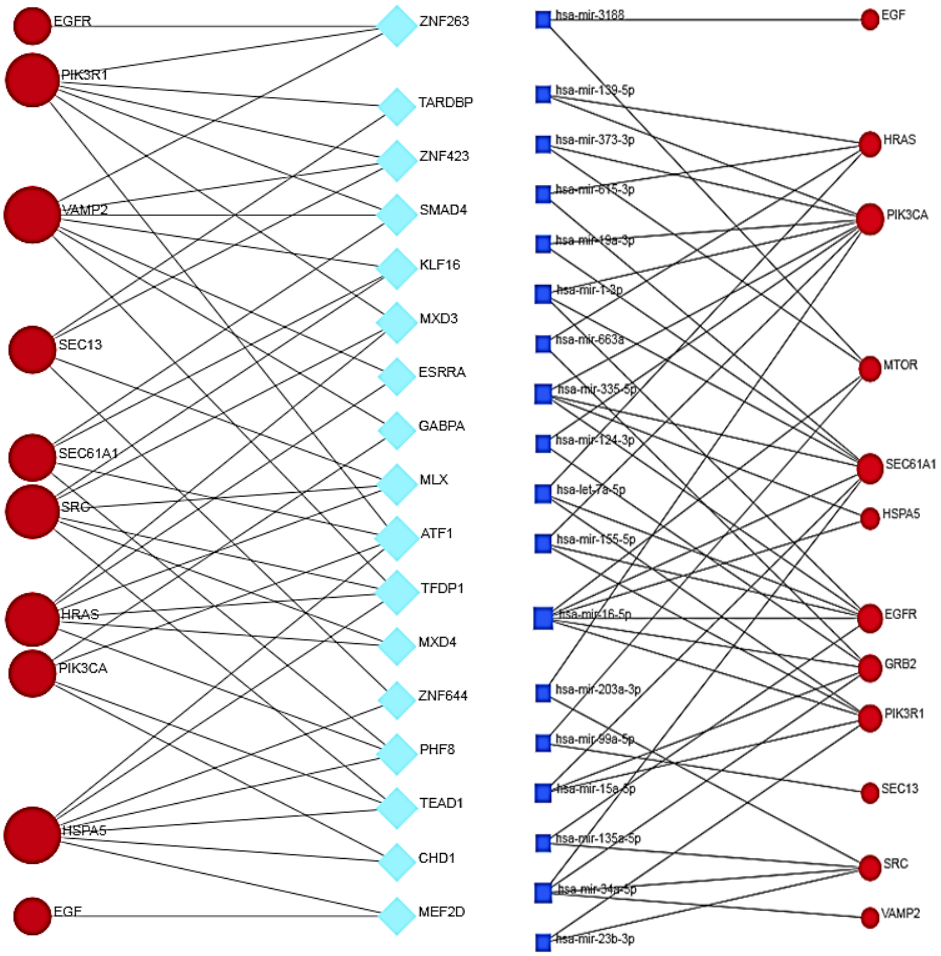


Figure 10. Transcriptional network (left) and post-transcriptional network (right) of interactions between genes (Hub and bottlenecks) and TFs and miRNAs, respectively. Red circles, genes; azure diamonds, TFs; blue rectangles, miRNAs. Rank of nodes in the networks is high as they undergo filtering based on degree and betweenness values. This is only a schematic view of the most significant molecules and their targets and where the size of the node is proportional to its rank.

 doi:10.20944/preprints202403.0473.v1

3.14. Co-Regulatory Network

Getting a co-regulated network requires integration of HUBs and bottlenecks with FTs and miRNAs. To determine the transcriptional regulatory relationships that these nodes may hold, we employed hub and bottleneck as enrichment seeds. This co-regulated network allowed us to pinpoint the 14 most reliable TFs and 2 miRNAs that were significantly associated with the expression of HUB and bottleneck genes.

The network (**Figure 11**) shows that among bottlenecks SEC13 is one of the most regulated genes. The protein encoded by this gene belongs to the SEC13 family of WD-repeat proteins and is a component of several important complexes. It is a component of the nuclear pore complex (NPC), which regulates transport between the nucleus and cytoplasm and has a direct role in regulating gene expression [104]. It is also a component of the COPII Coat Complex, where it plays a role in the formation of coated vesicles [105]. Four of the transcription factors that regulate SEC13 also regulate PIK3R1, the gene responsible for encoding Human_P85A, a protein that modulates glucose uptake in insulin-sensitive tissues by binding to activated Tyr kinases on the cellular membrane. Due to its inhibitory action, it appears to be a significant factor contributing to the hyperglycemia observed in covid patients. EGFR is also controlled by several TFs. The governance of each of these genes is multifaceted and bolstered by two miRNAs, specifically hsa-miR-576-5p and hsa-miR-1. The role of miRNA expression levels in disease processes and physiological development is significant, as changes in microRNA copy number or expression are closely associated with the onset of various human diseases [106]. miRNAs are present in a substantial number in humans [107].

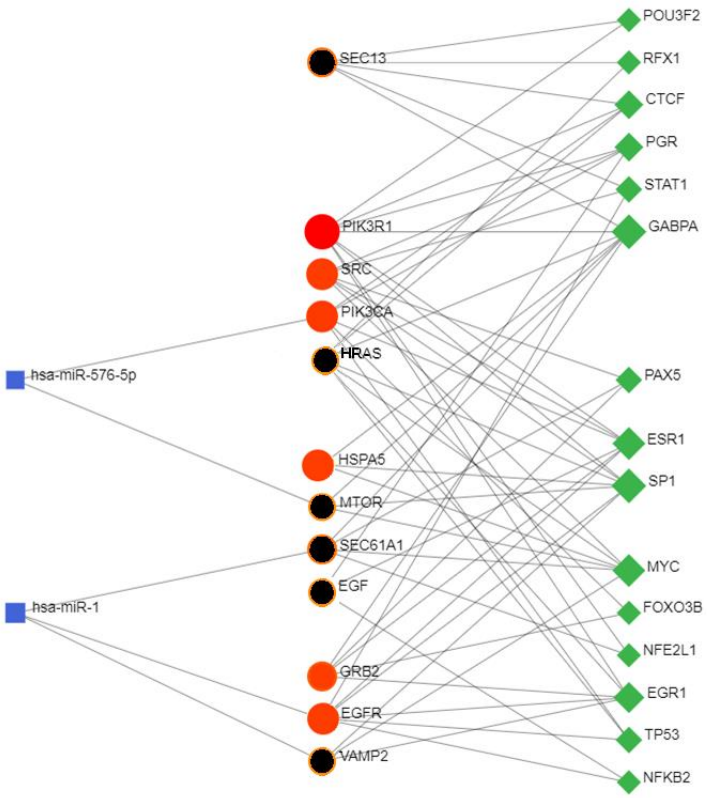


Figure 11. Integrated gene regulatory network associated with the dysregulated bottleneck and hub genes. Nodes: red orange circles, hubs; black circles, bottlenecks; green diamonds, TFs; blue rectangles, miRNAs. The figure shows the distribution of the potential gene–TF interactions (center and right side) and gene–miRNA interactions (center and left side). This is only a schematic view of the most significant molecules and their targets. We filtered the interacting network of miRNAs and TFs with betweenness centrality ≥ 100 and 45, respectively. **Figure 9S** displays the log-log graph, which confirms a scale-free distribution and shows some topological parameters.

 doi:10.20944/preprints202403.0473.v1

The correlation between miRNAs and human genes during SARS-CoV-2 infection is still an expanding research field with initial studies. Some preliminary evidence shows potential associations between miRNAs and genes that participate in the reaction to infection. It is essential to highlight that the analysis of this subject is still in progress. Despite this, miRNAs may be related to genes during SARS-CoV-2 infection to control inflammation. miRNA-155 [108] links the regulation of genes involved in inflammation, such as tumor necrosis factor alpha (TNF-α) and interleukin-6 (IL-6). According to previous research on COVID-19, miRNA-146a may be involved in regulating the innate immune response [109], and its upregulation may contribute to the dysregulation of inflammatory pathways. miRNAs might exert direct control over the replication of SARS2, as well as its capacity to infect host cells [110]. This could involve both the regulation of viral proteins and genes/proteins involved in human metabolism. Observations in cell lines and cancer patients led researchers to predict that miR-576-5p could down-regulate both PIK3CA and its mRNA [111]. Meanwhile, their target mRNAs were up-regulated. Hsa-miR-1 is believed to be linked to regulating human genes, especially in cancer patients [112]. Additionally, it has been noted that this specific miRNA also plays a role in the disturbance of glycemia for individuals with type 2 diabetes [113].

The co-regulatory network provides a better picture of metabolic events that the simple identification of a gene or protein in a metabolic pathway cannot give, even more so when we study the molecular mechanisms involved in a pathology. Merely asserting the involvement of a protein or gene in a pathological state without comprehending the coordinated activity of genes, miRNAs, TFs, mRNAs, and proteins may not always culminate in accurate inferences. Co-regulatory networks offer more decisive direction by elucidating the general coordination of the aforementioned actors, besides the appraisal of the pathological consequences of ORF7b.

3.15. Comparative Analysis of Negative Regulations According to the GO

Figure 12 shows the set of negative regulations vital for cellular diffusion represented by three transcriptional networks, which, upon comparison, exhibit remarkable similarities. In all three networks, EGFR, HRAS, HSPA5, PIK3CA, PIK3R1, and SRC are the genes involved in the negative control of programmed death. Their transcription at the individual gene level is negatively controlled through DNA-dependent transcription.

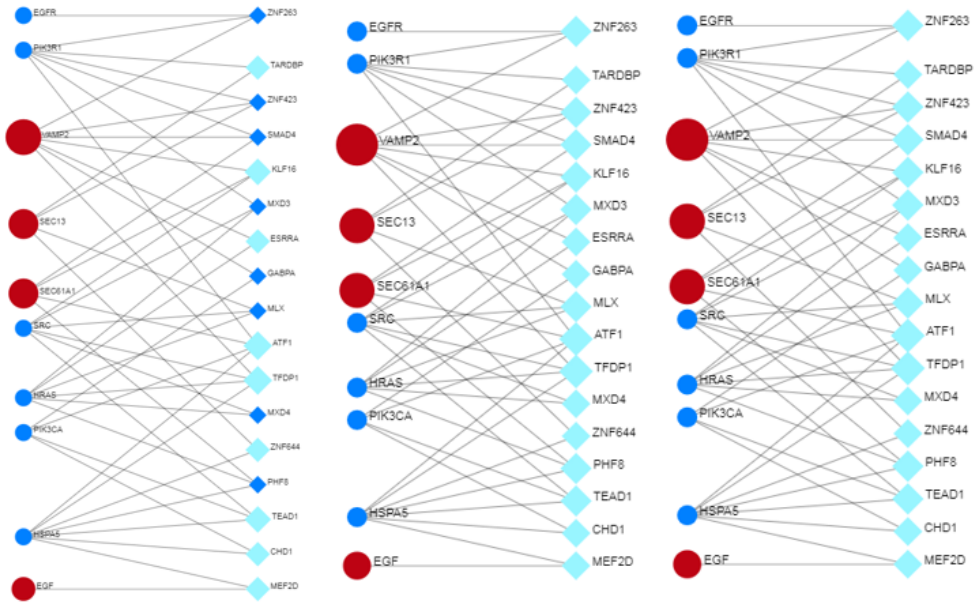


Figure 12. Comparison of three transcriptional networks related to negative metabolic controls because of ORF7b interference. GO analysis (genes in blue, bottlenecks in red). **Left side - Negative regulation of transcription, DNA_dependent (p <1,56e-4)** (EGFR, HRAS, HSPA5, PIK3CA, PIK3R1, SRC, and ZNF263, ZNF423, SMAD4, MXD3, GABPA, MLX, MXD4, PHF8) **Middle** - **Negative**

 doi:10.20944/preprints202403.0473.v1

regulation of apoptotic process (p <7.58e-4) (EGFR, HRAS, HSPA5, PIK3CA, PIK3R1, SRC) **Right side - Negative regulation of programmed cell death (p <8.86e-4)** (EGFR, HRAS, HSPA5, PIK3CA, PIK3R1, SRC)

Below is a brief illustration of the most intriguing transcription factors found in the networks. ZNF423 and ZNF263 (Zinc Finger Protein 423 and 263) can act as both transcriptional repressors and activator by binding to DNA, where ZNF423 plays a central role. MXD4 (Max-interacting transcriptional repressor MAD4) is a protein that in humans is encoded by the MXD4 gene. PHF8 (Histone lysine demethylase PHF8) is a transcription activator which acts on the epigenetically methylated Histone 3 but is a repressor for the methylated histone 4. Acts as a coactivator of rDNA transcription, by activating polymerase I (pol I) mediated transcription of rRNA genes and playing a role in the cell cycle. However, its role remains still unsure in vivo. GABPA (GA Binding Protein Transcription Factor Subunit Alpha) is a transcription factor interacting with purine rich repeats (GA repeats), so positively regulating the transcription of transcriptional repressor RHIT and of ZNF family such as ZNF205. MLX (MAX Dimerization Protein MLX), its decoded product (Max-like protein X) forms many sequence-specific DNA-binding protein complexes with various proteins. These complexes act as transcriptional repressors. Plays a peculiar role as a transcriptional activator of glycolytic target genes, thus it is involved in glucose-responsive gene regulation. Here, we have another pro-glycemic effect, common to covid patients.

While the **Figure 13** shows the relationships between genes and miRNAs involved in blocking programmed death at the post-transcriptional level the **Figure 14** shows its co-regulated network where we find the occurrence of Myc and TP53, two well-known transcription factors. MYC, (MYC Proto-Oncogene or BHLH Transcription Factor, which codes for P01106 · MYC_HUMAN) is involved in many diseases (114). The Gene Ontology (GO) annotations that concern MYC comprise DNA-binding transcription factor activity, and the ability to function with TAF6L to activate target gene expression through RNA polymerase II cis-regulatory region sequence-specific DNA binding.

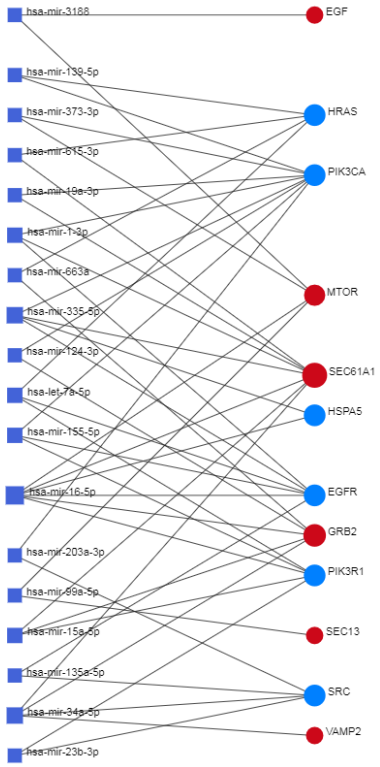


Figure 13. – Post-transcriptional networks related to negative metabolic controls because of ORF7b interference. GO analysis for **Negative regulation of programmed cell death (p <8.28e-6)** (EGFR, HRAS, PIK3R1, HSPA5, PIK3CA, SRC).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 March 2024 doi:10.20944/preprints202403.0473.v1

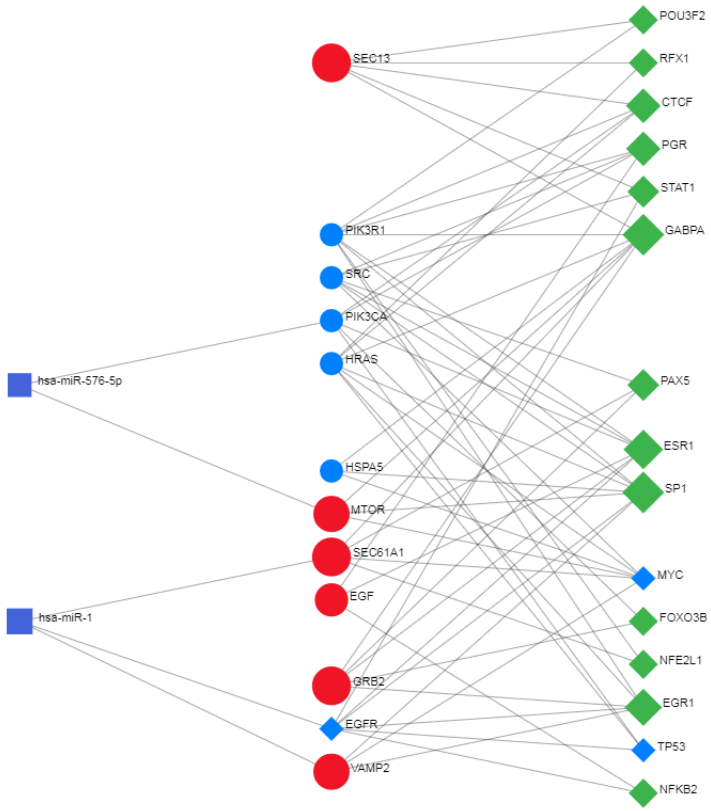


Figure 14. Co-regulated network related to negative metabolic controls because of ORF7b interference. GO analysis for **Negative regulation of programmed cell death** (p<1.82e-5). (TFs: TP53, MYC; Genes: SRC, EGFR, HRAS, PIK3R1, HSPA5, PIK3CA; miRNA: has-miR-1 and has-miR-576-5p).

TP53, also known as Tumor Protein 53 and encoding for P04637, Cellular tumor antigen p53, acts as a tumor suppressor, in response to cellular stresses to regulate the expression of target genes [115]. However, in specific metabolic contexts, it can induce cell cycle arrest, apoptosis, and changes in metabolism [116]. As a matter of fact, it has been discovered that SARS-CoV-2 infection leads to the stabilization of TP53 on chromatin [117], contributing to a robust host cytopathic effect. Modification of chromatin accessibility, cellular senescence, and inflammatory cytokine release through TP53 is brought about by the involvement of this protein in various SARS-CoV-2 spike variant-induced syncytia formations. The protein appears to have a role in inflammation associated with cellular senescence [117]. In addition, TP53 was discovered to be implicated in IFN-γ-mediated signaling, apoptosis, and proteasomal degradation of CD4 T cells [118]. However, uncertainties regarding the functionality of miRNAs persist because of technical difficulties and the considerable number of miRNAs that are still subject to systematic profiling [107]. Because of their low intrinsic stability and RNAses [119], they are susceptible to degradation, and laboratory manipulations can have questionable effects on their measurements [120,121].

TFs are well-established proteins with reliable experimentally derived results, although miRNAs remain somewhat enigmatic. TFs are proteins that control the rate of transcription of genetic information from DNA to mRNA binding to DNA. Thus, their function is to regulate, switching on and off, genes. This functional activity is to address the gene expression to the exact target cells at the right time and in the right amount. Groups of TFs function in a coordinated fashion to direct cell division, cell growth, and cell death. TFs work alone or with other proteins in a complex, by promoting (as an activator), or blocking (as a repressor) the recruitment of RNA polymerase to specific genes.

We have examined the various correlations between miRNAs, TFs and the components of the compact hub-and-spoke architectural system of the PPI network, getting information on the fundamental co-regulations operated by some TFs and miRNAs. These findings suggest that

 doi:10.20944/preprints202403.0473.v1

crosstalk motifs, comprising the direct and non-shared relationships between regulators and their target genes, can have downstream effects on diverse biological processes, in line with the features already highlighted in the interactome's analysis network. This analysis amplifies and substantiates our findings and deductions from the interactomic analysis. Our result, however, has limitations. The human genome contains thousands of coding and non-coding RNA genes. These genes express differentially, in diverse locations, at distinct times during normal homeostasis, or in response to environmental cues. The differential expression extends to TFs and miRNAs, too. The regulation of genes is specific to certain conditions and changes over time, meaning our findings only provide a static view of the molecular mechanisms affected by ORF7b. While our conclusions are valid, we can only show the presumed targets, not how they dynamically work.

4. Discussion

The guiding principle that underpins this research is that SARS-CoV-2 infection leads to changes in the deep metabolic activities of infected cells to favor the acquisition and maintenance of viral strategies, compared to normal cells. The virus causes a reprogramming of cellular metabolism by its proteins. The expression "metabolic reprogramming" refers to the recognition of normal metabolic pathways which are modified by viral proteins, when compared to those in normal tissue. This point is significant because the analysis of "metabolic normality" is often overlooked. Our training in cancer has taught us to search for mutations that can modify signaling processes. In viral infections, mutations are absent, as viruses achieve the same aim by up- or down-regulating normal signaling pathways or other metabolic processes.

Our results reveal the functional impact of the accessory protein ORF7b in SARS-CoV-2 infection and identify molecules that control metabolic processes dysregulated by this viral protein. Among the many functional activities highlighted, we focused on those that promote the spread of infected cells in the organism.

The release of virions into the extracellular space is a common event among numerous viruses, which has stimulated the study of virus egress/entry biology. Although some viruses spread through the shedding of infected cells [180,181], this is an understudied topic. Recently, several authors [182,183] have reported evidence by antibody experiments, that SARS-CoV-2 could spread through cell-to-cell transmission. However, no one has studied or hypothesized any related molecular mechanism. In this article, we confirm those authors' hypotheses and describe the deep molecular mechanism that underlies this feature of SARS-CoV-2. This discovery contributes to our understanding of the human immune response to the attack of this virus because cell-to-cell transmission is an effective means by which viruses evade host immunity.

It is still to be considered that our data on spread in some sense support the theory on the evolution of virulence which assumes that high growth rates of pathogens should both increase transmission between hosts and increase disease induced morbidity or mortality [186,187]. The spread of infected cells could fit into this logic, but the theory also suggests that through viral “tolerance”, virulence is mitigated without reducing viral load [187,188]. This also dictates that the host should allow the selection of the pathogen with a higher growth rate to gain a gain in transmission between hosts but without causing harm to the original host [188–190]. Today's clinical data tells us that the virulence of Covid-19 is decreasing without any type of specific intervention. Our data does not explain the effects of diffusion on the virulence, but it paves the way for experimental designs with greater awareness of what actually happens.

We have analyzed by interactomics only functional and physical correlations between ORF7b and the entire human proteome determined by experiments. To obtain reliable interactomics results, we extracted from the set of interactors only those that were characterized by high significance. The investigation showed that the virus achieves its strategic goals by interacting with metabolic processes controlled by human proteins such as EGFR, SRC, HSPA5, MTOR, SEC13, SEC61A1, VAMP2, PIK3R1, PIK3CA, GRB2, and HRAS, which are important for human metabolism because they are high-ranking, HUB and bottleneck proteins. Through a series of analyses using transcriptional co-regulation networks, we have also validated our results by identifying regulatory

 doi:10.20944/preprints202403.0473.v1

actions conducted by transcription factors and miRNAs on genes that code for the previously identified key proteins.

Viruses do not perform metabolic processes but know how to interact with them to their own advantage. Although various attempts have been made to identify metabolic pathways and nodes under the control of the virus, to our knowledge, this is the first wide-ranging interactome map identified for a specific protein of SARS-CoV-2. We have identified some metabolic pathways under the control of ORF7b, but still, we have a limited knowledge of the comprehensive set of viral proteins involved, and by which specific mechanisms. Despite that, several authors have hypothesized some functional activities of ORF7b in the infected cell and its synergism with other viral proteins, but no one has attempted to study in depth the molecular and functional interactions within human metabolism implemented by ORF7b. In particular, they identified its involvement in SNARE-driven vesicular transport, exocytic processes, ERBB signaling, but without a functional characterization that identifies the actual role of ORF7b, possibly in synergy with other viral proteins [122]. There have been other studies that have endeavored to juxtapose the mechanisms of diseases between SARS1 and SARS2 [123,124], but none of them have deciphered any common molecular mechanism. Both viruses lead to acute respiratory distress, but many phenomenological observations show differences. One study predicted that SARS-CoV-2 induces a systemic disease, which, unlike SARS1, damages various organs in the body, such as the heart, kidney, and brain [125,126]. These results suggest that the two viruses use different molecular mechanisms but also that we do not know which mechanisms they use. Out of curiosity, searching PubMed for "differences in molecular mechanisms of SARS-CoV-1 and SARS-CoV-2" or "molecular mechanisms of SARS-CoV-1 and SARS-C0V-2 (or similar terms), yielded no results.The continuous work conducted by the curators of BioGRID, in selecting and evaluating the statistical significance of each single experimentally characterized interaction between the viral proteins and the human proteome, has allowed us to design this study with the methods of Interactomics. Direct knowledge of the deep molecular mechanisms implemented by individual viral proteins is essential because only through this knowledge will we be able to design specific and effective antiviral drugs. A study at a deep molecular level, a research area still rather obscure in its modes of action in space and time, is an important approach if aimed at identifying those human proteins which, in viral infection, play crucial roles such as hub nodes or as a bottleneck. These proteins represent the crossroads of multiple biological activities and, therefore, are the best targets for disease control.

ORF7b is a tiny viral protein of 43 amino acids, a macro-polyanion with a net charge of –4 at neutral pH (four negative residues and no positive charge); the central part from 9 to 29 is helical, and the protein surface is negative [6]. This protein does not appear to operate on its own (see **TABLE IX**). What emerges from this study is the precise interference of ORF7b into various molecular mechanisms at the basis of our metabolism. ORF7b showed mostly diverse behaviors, in terms of localization, membrane recruitment, and metabolic dynamics. The results show its important role in conditioning cellular transport processes as well in some important signaling pathways (see **Table III**). The topological characteristics of this interactome reveal a group of proteins with structural and functional properties that are consistently implicated in multiple metabolic activities, some of which are dysregulated by ORF7b's action. These proteins are characterized by their high degree of functional relationships and by their high ability to regulate a multitude of significant metabolic and signaling pathways. The interactome shows certain metabolic modules that perform necessary functional activities for normal cellular metabolism. A large central core (GCC) comprising two closely connected clusters was identified through cluster analysis as the primary functional location of these proteins. The high number of tight connections favors a high metabolic rate, which accelerates any functional activity.

The activity of these proteins extends to very different places in the cell (see **Table 4**) according to a hub-and-spoke topological model and potentially also to those tissues that have the molecular characteristics suitable for the entry of the virus (see **Table 6**). All this suggests that ORF7b must have a remarkable ability to interact with different molecular partners, such as to allow it to operate practically everywhere, at the membrane level and in the cytoplasm. Indeed, the list of its main

 doi:10.20944/preprints202403.0473.v1

molecular interactors shows both membrane proteins and cytoplasmic proteins. Some authors have hypothesized a role for ORF7b as an intrinsic single-span membrane protein, in analogy with the 44 amino acid homolog ORF7b of SARS [6]. This hypothesis is rather restrictive considering the wide spectrum of functions in which this protein is involved and the spatio-temporal characteristics that a biological object of this type must possess in order to be involved in the various intracellular transport processes (see GO:0006810, p-value 3.04e-67); or even having also to guide and regulate target localization (see GO:0008104 and GO:0045184, p-values 1.4e-58 and 2.85e-58, respectively). But, at the same time this protein must also have the characteristic to interact with different membrane systems (see GOCC:0016020, GOCC:0031090, GOCC:0031982 or GOCC:0098588, with p-values of 2.5e-92, 2.07e-77, 5.13e-62 and 3.17e- 58, respectively) and to interfere with metabolic signaling paths (see GO:0007169, GO:0007167; HAS-9006934, HAS-1227986, or HAS-6811558, with p-values of 1.23e-66, 7.95e-59, 4.44e-84, 2.43e-30, and 5.16e-24, respectively).

The regulated functional re-localization seems one of the most important characteristics of this protein [127]. ORF7b shows coherent functional solutions with viable biochemical functional models. The closest class of proteins possessing these types of broad properties is called the “Peripheral Membrane Proteins”, a class of proteins that live at the membrane interface [128,129]. In 2002, Felix Goñi [130] introduced the concept of "non-permanent membrane proteins", to encompass the wide variety of proteins that are not found in a stable membrane-bound form under physiological conditions, but interact with the membrane in certain phases of their specific course of action. Despite the fundamental biological meaning of these proteins, an experimental characterization of their structure has always been vague because attempts at structure prediction often fail. Therefore, this protein class has a poor representation of its 3D structures within the PDB because they are difficult to study [131]. Wanting to represent them in a few words, they are soluble proteins that bind transiently to the surface of biological membranes or even to proteins on the outer side of the membrane, where they perform their functions. The reversible attachment of proteins to biological membranes shows how they can regulate cell signaling and many other important cellular events, through a variety of mechanisms [132,133]. Thus, the behavior of peripheral proteins, reversibly associated with the lipid bilayer [132,134], may also explain the behavior of ORF7b, coherently with its structural/functional properties. Therefore, this protein appears as a very reliable member of the class of "non-permanent membrane proteins" [135].

Recent molecular dynamics simulations experiments provided molecular insights of the protein's dimerization [191]. This study shows different dimerization models, parallel and antiparallel. Among the various structures modeled, the authors suggest the possibility that the parallel dimer may operate docked to the membrane, from 7 to 30, and floating in the cytoplasm from 31 to 43. According to authors, while simulations support the homodimerization of ORF7b, the analysis of genetic mutations of orf7b during the evolution of the pandemic suggests an unstable dimerization when associated with the regulation of IFN production, the apparent function attributed to this protein. They conclude that the lack of detailed structural information on lateral protein-protein associations hinders a thorough evaluation of packing, which means that there is not yet sufficient detail to define consistent structure-function relationships. This information adds to the previous considerations but every hypothesis made still remains valid.

Like other viruses, SARS-CoV-2 can cause reinfection/reactivation and persistent infection, as supported by several experimental studies [136]. SARS-CoV-2 has the potential to activate or modulate oncogenic cancer-promoted pathways, leading to chronic low-grade inflammation and tissue damage, according to growing evidence [137]. Several authors perceive oncogenesis as a potential long-term effect of SARS-CoV-2 infection, which could lead to the onset of cancer by inhibiting tumor suppressor genes [138]. The utilization of similar tactics as EBV or HSV1 by SARS-CoV-2 to manipulate p53 is clear, as the virus takes over the protein using viral antigens, which lead to p53 deterioration [139]. By deactivating both external and internal apoptotic pathways of host cells, SARS-CoV-2 may spread like cancer cells [140,141]. Our results suggest that the cancer-like effects of SARS-CoV-2 result from the virus capability to spread infected cells through the action of its proteins, mimicking cancer and its metastasis. The lack of adequate understanding of the mechanisms that

 doi:10.20944/preprints202403.0473.v1

govern the progression of the virus after the release of infected cells makes it impossible to make accurate predictions about the long-term implications of the long-covid.

However, we should make a last consideration given the recent advances in our understanding of the N protein of SARS-CoV-2. Phosphorylation of the central disordered region of the N protein forms dynamic, liquid-like condensates that control also the viral genome transcription [142]. N protein contains three dynamic disordered regions that house putative transiently helical binding motifs and the protein undergoes liquid-liquid phase separation [142,143] thus phosphorylation regulates the accessibility and assembly of N protein to bio-condensate [144]. Another critical function of N is to encapsulate the viral genome of ssRNA to evade immune detection and protect viral RNA from degradation by host factors [144–147].

Viral proteins form condensates for their molecular strategies, such as infection and signaling transduction [148,149]. Viruses regularly execute their molecular tactics in specific parts of cells. For instance, we should consider how phase separation in cell compartments affects important processes like viral transcription or viral spread [148,149]. The fact that ORF7b interacts with N (**TABLE IX**) supports the involvement of ORF7b in viral diffusion phenomena, with mechanisms of alteration of the cytoskeleton, but, perhaps, also with more complex mechanisms involving liquid droplets. After all, phase separation is one of the basic molecular processes that govern multiple cellular activities, such as cancer progression, gene expression, and signaling transduction [150].

In SARS-CoV-2, the properties of the liquid-like condensate that forms phase-separated compartments without a membrane and the transient nature of interactions within them, are determined by the interaction of N protein with viral RNA because of its intrinsic disorder properties [151]. The threshold for phase separation decreases as the number of interacting sites of a molecule increases. This multivalency comes from structural domains, where each domain contributes to binding [151]. Intrinsically disordered-regions (IDRs) often participate in phase separation, as they might provide a source of multivalency. In fact, the low affinity of their individual interactions can enable liquid-like properties [151].

We are studying SARS-CoV-2-host interactions in a simplistic context because crucial information is missing from the databases. For example, there have been few studies evaluating virus-host molecular interactions considering the range of post-translational modifications. Without the enormous potential of the biological role of post-translational modifications, we run serious risks of having distorted information on the biology of this virus. Researchers have shown crucial roles of phosphorylation and ubiquitylation in other systems, but have not yet identified the corresponding proteoforms in SARS-CoV-2-host interactions.

However, many of the high-ranking proteins we have studied and selected show that they have all the characteristics necessary to act even through forms of bio-condensates. Therefore, we cannot exclude that, together with the co-regulation that we have highlighted, there may also be a further form of regulatory activity exerted by the liquid-like condensates. We cannot exclude it, considering their well-established presence in cells and the important roles they play.

5. Conclusions

The proposed model, although created on the most robust basis possible in consideration of our current knowledge on the interactions of ORF7b and other SARS-CoV-2 proteins with the human proteome (see “Robustness of the study” in Supplements), will definitely be worth seeing again, supported by more precise knowledge on transcriptional modifications and the spatio-temporal characteristics of its proteins, and by the role of bio-condensates. Deep biological aspects are still very little known, and often overlooked. Without this, we will continue to have inconsistent flat views of metabolism and viral action.

Notably, our work offers a mechanistic hypothesis to explain aspects of the virulence of SARS-CoV-2. More generally, demonstrates key differences in using the mesoscopic approach of Systems Biology compared to symptom-based macroscopic approaches, which tells us very little about what might be happening at deep metabolic levels in the human body. Obviously, as long as it is based exclusively on omics data, experimental and significant, because this is the real limit. Our

interactomics framework indeed offers a series of testable questions and predictions, drawable to stimulate future work, such as comparing deep mechanisms of virulence evolution in diverse infection stages. Understanding the molecular mechanisms that select the evolution of viral traits in the human host should allow us to better predict and combat the virulence of probable future threats, and also to understand the most suitable targets for designing a drug.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, G.C.; G.F and G.M. carried out genotype analyses; G.C. carried out interactomic analyses; G.C., G.M., and G.F. contributed to critical analysis of the data; writing—original draft preparation, G.C., G.M., and G.F.; writing—article and editing, G.C., G.M.,; funding acquisition, G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a Grant to Giovanna Fusco, Ministero della Salute, number C73C18000160005, (invoice addressed to Giovanna Fusco, Istituto Zooprofilattico Sperimentale del Mezzogiorno, via Salute, 2 - 80055 Portici (Napoli) Italy").

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

References

1. Bar-On Y.M., et al., SARS-CoV-2 [COVID-19] by the numbers. eLife [2020]; 9:1-15; e57309. DOI: https://doi.org/10.7554/eLife.57309.
2. Sender R., et al., The total number and mass of SARS-CoV-2 virions. PNAS. [2021] volume 118, 25, pages = {e2024815118}, doi = {10.1073/pnas.2024815118}, https://www.pnas.org/doi/abs/10.1073/pnas.2024815118.
3. Milo R. What is the total number of protein molecules per cell volume? A call to rethink some published values. Bioessays. 775 [2013] Dec;35[12]:1050-5. doi: 10.1002/bies.201300066. Epub 2013 Sep. PMID: 24114984; PMCID: PMC3910158. 776
4. Andreas-David Brunner et al., Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. Mol Syst Biol [2022] 18:e10798 https://doi.org/10.15252/msb.202110798 . 780
5. Brandon Ho, et al., Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. [2018], Cell Systems 6, 1–14. https://doi.org/10.1016/j.cels.2017.12.004
6. Colonna, G. A Tiny Viral Protein, SARS-CoV-2-ORF7b: Structural Features. *Preprints* **2023**, 2023040522. https://doi.org/10.20944/preprints202304.0522.v1
7. Zhang, J., et al., . A systemic and molecular study of subcellular localization of SARS-CoV-2 proteins. *Sig Transduct Target Ther* **5**, 269 (2020). https://doi.org/10.1038/s41392-020-00372-8
8. Oughtred R, et al., The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. [2020] Oct 18.
9. S. Boccaletti , et al., Hwang Complex networks: structure and dynamics Phys Rep, 424 (4) (2006), pp. 175-308
10. Chris Stark, et al., BioGRID: a general repository for interaction datasets, *Nucleic Acids Research*, Volume 34, Issue suppl_1, 1 January 2006, Pages D535–D539, https://doi.org/10.1093/nar/gkj109
11. Chen Z, et al., Comprehensive analysis of the host-virus interactome of SARS-CoV-2 (2021 Pre-Publication) Status: Preliminary Report
12. Stukalov A, Girault V, et al., Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. (2021) Nature; Jun;594(7862):246-252. doi: 10.1038/s41586-021-03493-4. Epub 2021 Apr 12. Pubmed:33845483

13. Estelle M.N. Laurent, Yorgos Sofianatos, et al., Global BioID-based SARS-CoV-2 proteins proximal interactome unveils novel ties between viral polypeptides and host factors involved in multiple COVID19-associated mechanisms (2020 Pre-Publication). Status: Preliminary Report
14. Samavarchi-Tehrani P, et al., A SARS-CoV-2 - host proximity interactome (2020 Pre-Publication) Status: Preliminary Report.
15. St-Germain JR, Astori A, et al., A SARS-CoV-2 BioID-based virus-host membrane protein interactome and virus peptide compendium: new proteomics resources for COVID-19 research (2020 Pre-Publication) Status: Preliminary Report
16. Bamberger TC, Pankow S, et al., The Host Interactome of Spike Expands the Tropism of SARS-CoV-2 (2021 Pre-Publication). Status: Preliminary Report.
17. Liu X, Huuskonen S, et al., SARS-CoV-2-host proteome interactions for antiviral drug discovery. (2020) Mol Syst Biol. Nov;17(11):e10396. PMID: 34709727 PMCID: PMC8552907 DOI: 10.15252/msb.202110396.
18. Snider J, et al., Fundamentals of protein interaction network mapping. Mol Syst Biol. 2015 Dec 17;11(12):848. doi: 10.15252/msb.20156351. PMID: 26681426; PMCID: PMC4704491.
19. O'Connor, D.J. (1983). Emergent Properties. In: van der Merwe, A. (eds) Old and New Questions in Physics, Cosmology, Philosophy, and Theoretical Biology. Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-8830-2_48.
20. Boojari, M.A. Investigating the Evolution and Development of Biological Systems from the Perspective of Thermo-Kinetics and Systems Theory. *Orig Life Evol Biosph* **50**, 121–143 (2020). https://doi.org/10.1007/s11084-020-09601-0
21. Szklarczyk D, et al., STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015 Jan; 43(Database issue): D447-52. doi:10.1093/nar/gku1003. Epub 2014 Oct 28. PMID: 25352553; PMCID: PMC4383874.
22. Machicao J, et al., On the Use of Topological Features of Metabolic Networks for the Classification of Cancer Samples. Curr Genomics. 2021 Feb;22(2):88-97. doi: 10.2174/1389202922666210301084151. PMID: 34220296; PMCID: PMC8188584.
23. Xueming Liu, et al., Controllability of giant connected components in a directed network. PHYSICAL REVIEW E 95, 042318 (2017), DOI: https://doi.org/10.1103/PhysRevE.95.042318
24. Wagner, How the global structure of protein interaction networks evolves. Proceedings of the Royal Society of London. Series B, (2003), 270,457–466. DOI:10.1098/rspb.2002.2269
25. Shannon P, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769.
26. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. Bioinformatics. 2008 Jan 15;24(2):282-4. doi: 10.1093/bioinformatics/btm554. Epub 2007 Nov 15. PMID: 18006545.
27. Doncheva, N., *et al.* Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protoc* **7**, 670–685 (2012). https://doi.org/10.1038/nprot.2012.004
28. Dong, J., Horvath, S. Understanding network concepts in modules. *BMC Syst Biol* **1**, 24 (2007). https://doi.org/10.1186/1752-0509-1-24
29. Kumar M, Saini S, Gayen K. Elementary mode analysis reveals that Clostridium acetobutylicum modulates its metabolic strategy under external stress. Mol Biosyst. 2014 Aug;10(8):2090-105. doi: 10.1039/c4mb00126e. PMID: 24852622.

30. Barabási, A.L., Oltvai, Z.N.: Network biology: understanding the cell's functional organization. Nat Rev Genet **5** (2004) 101-113
31. A.Wagner, How the global structure of protein interaction networks evolves. Proc.R.Soc.Lond.B (2003), 270, 457–466 457 DOI10.1098/rspb.2002.2269
32. M. Gerlach, E. G. Altmann, Testing statistical laws in complex systems. Phys. Rev. Lett. 122, 168301 (2019).
33. Jose C Nacher and Tatsuya Akutsu, Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. 2012 New J. Phys. 14 073005. DOI 10.1088/1367-2630/14/7/073005
34. Winterbach W, et al., Topology of molecular interaction networks. BMC Syst Biol. 2013 Sep 16;7:90. doi: 10.1186/1752-0509-7-90. PMID: 24041013; PMCID: PMC4231395.
35. Yoon TY, Munson M. SNARE complex assembly and disassembly. Curr Biol. 2018 Apr 23;28(8):R397-R401. doi: 10.1016/j.cub.2018.01.005. PMID: 29689222.
36. Kádková A, Radecke J, Sørensen JB. The SNAP-25 Protein Family. Neuroscience. 2019 Nov 10;420:50-71. doi: 10.1016/j.neuroscience.2018.09.020. Epub 2018 Sep 27. PMID: 30267828.
37. Grabowski P, et al.,. "Expression of neuroendocrine markers: a signature of human undifferentiated carcinoma of the colon and rectum". Virchows Archiv. (2022) **441** (3): 256–263. doi:10.1007/s00428-002-0650-9. PMID 12242522. S2CID 21987523.
38. Xi Z, et al., "Association of Alpha-Soluble NSF Attachment Protein with Epileptic Seizure". Journal of Molecular Neuroscience. (2015) **57** (3): 417–425. doi:10.1007/s12031-015-0596-4. PMID 26156199. S2CID 16026967.
39. Morton AJ, Faull RL, Edwardson JM. "Abnormalities in the synaptic vesicle fusion machinery in Huntington's disease". Brain Research Bulletin. (2001) **56** (2): 111–117. doi:10.1016/S0361-9230(01)00611-6. PMID 11704347. S2CID 35996576.
40. Nalbandian A, et al., Post-acute COVID-19 syndrome. Nat Med. 2021 Apr;27(4):601-615. doi: 10.1038/s41591-021-01283-z. Epub 2021 Mar 22. PMID: 33753937; PMCID: PMC8893149.
41. Silva Andrade B, et al., Long-COVID and Post-COVID Health Complications: An Up-to-Date Review on Clinical Conditions and Their Possible Molecular Mechanisms. Viruses. 2021 Apr 18;13(4):700. doi: 10.3390/v13040700. PMID: 33919537; PMCID: PMC8072585.
42. Vadivalagan C, et al., Exosomal mediated signal transduction through artificial microRNA (amiRNA): A potential target for inhibition of SARS-CoV-2. Cell Signal. 2022 Jul;95:110334. doi: 10.1016/j.cellsig.2022.110334. Epub 2022 Apr 21. PMID: 35461900; PMCID: PMC9022400.
43. Krachmarova E, et al., Insights into the SARS-CoV-2 ORF6 Mechanism of Action. Int J Mol Sci. 2023 Jul 18;24(14):11589. doi: 10.3390/ijms241411589. PMID: 37511350; PMCID: PMC10380535.
44. Rozen EJ, Shohet JM. Systematic review of the receptor tyrosine kinase superfamily in neuroblastoma pathophysiology. Cancer Metastasis Rev. 2022 Mar;41(1):33-52. doi: 10.1007/s10555-021-10001-7. Epub 2021 Oct 30. PMID: 34716856; PMCID: PMC8924100.
45. Wang Z. ErbB Receptors and Cancer. Methods Mol Biol. 2017;1652:3-35. doi: 10.1007/978-1-4939-7219-7_1. PMID: 28791631.
46. Hadid T, et al.,. Coagulation and anticoagulation in COVID-19. Blood Rev. 2021 May;47:100761. doi: 10.1016/j.blre.2020.100761. Epub 2020 Oct 8. PMID: 33067035; PMCID: PMC7543932.
47. Khunti K, et al., COVID-19, Hyperglycemia, and New-Onset Diabetes. Diabetes Care. 2021 Dec;44(12):2645-2655. doi: 10.2337/dc21-1318. Epub 2021 Oct 8. PMID: 34625431; PMCID: PMC8669536.

48. Loughrey D, Dahlman JE. Non-liver mRNA Delivery. Acc Chem Res. 2022 Jan 4;55(1):13-23. doi: 10.1021/acs.accounts.1c00601. Epub 2021 Dec 3. PMID: 34859663.
49. Peiris S, et al., Pathological findings in organs and tissues of patients with COVID-19: A systematic review. PLoS One. 2021 Apr 28;16(4):e0250708. doi: 10.1371/journal.pone.0250708. PMID: 33909679; PMCID: PMC8081217.
50. Koçak Tufan Z, et al., COVID-19 and Sepsis. Turk J Med Sci. 2021 Dec 17;51(SI-1):3301-3311. doi: 10.3906/sag-2108-239. PMID: 34590796; PMCID: PMC8771020.
51. Kopańska M, et al., Effects of SARS-CoV-2 Inflammation on Selected Organ Systems of the Human Body. Int J Mol Sci. 2022 Apr 10;23(8):4178. doi: 10.3390/ijms23084178. PMID: 35456997; PMCID: PMC9025828.
52. 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.
53. Cifuentes-Munoz N, et al., Viral cell-to-cell spread: Conventional and non-conventional ways. Adv Virus Res. 2020;108:85-125. doi: 10.1016/bs.aivir.2020.09.002. Epub 2020 Sep 29. PMID: 33837723; PMCID: PMC7522014.
54. Tiwari V, et al., Role of Tunneling Nanotubes in Viral Infection, Neurodegenerative Disease, and Cancer. Front Immunol. 2021 Jun 14;12:680891. doi: 10.3389/fimmu.2021.680891. PMID: 34194434; PMCID: PMC8236699.
55. Ermagun, A., Tajik, N. & Mahmassani, H. Uncertainty in vulnerability of networks under attack. *Sci Rep* **13**, 3179 (2023). https://doi.org/10.1038/s41598-023-29899-w
56. Masoomy, H. and al., Relation between the degree and betweenness centrality distribution in complex networks. {Phys. Rev. E, vol 107, issue 4, pages 044303, (2023), American Physical Society, doi = 10.1103/PhysRevE.107.044303, https://link.aps.org/doi/10.1103/PhysRevE.107.044303
57. Vistain LF, Tay S. Single-Cell Proteomics. Trends Biochem Sci. 2021 Aug;46(8):661-672. doi: 10.1016/j.tibs.2021.01.013. Epub 2021 Feb 27. PMID: 33653632.
58. Rosa SS, et al., mRNA vaccines manufacturing: Challenges and bottlenecks. Vaccine. 2021 Apr 15;39(16):2190-2200. doi: 10.1016/j.vaccine.2021.03.038. Epub 2021 Mar 24. PMID: 33771389; PMCID: PMC7987532.
59. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol. 2007 Apr 20;3(4):e59. doi: 10.1371/journal.pcbi.0030059. Epub 2007 Feb 14. PMID: 17447836; PMCID: PMC1853125.
60. Di Silvestre D, et al., Network Topological Analysis for the Identification of Novel Hubs in Plant Nutrition. Front Plant Sci. 2021 Feb 10;12:629013. doi: 10.3389/fpls.2021.629013. PMID: 33679842; PMCID: PMC7928335.
61. Kei Yura, et al., Alternative splicing in human transcriptome: Functional and structural influence on proteins, Gene, Vol 380, Issue 2, 2006, Pages 63-71, ISSN 0378-1119, https://doi.org/10.1016/j.gene.2006.05.015.
62. Chandramohan Nithya, et al., Dissection of hubs and bottlenecks in a protein-protein interaction network, Computational Biology and Chemistry, Vol 102, 2023, 107802, ISSN 1476-9271, https://doi.org/10.1016/j.compbiolchem.2022.107802.
63. Manjari Kiran et al., Global versus Local Hubs in Human Protein–Protein Interaction Network, J. Proteome Res. 2013, 12, 12, 5436–5446. https://doi.org/10.1021/pr4002788
64. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37, 215–246 (2008).

65. Lazaros K.Gallos, et al., Stability and Topology of Scale-Free Networks under Attack and Defense Strategies. Physical Review Letters. **94**, 188701(1-4) (2005). DOI:10.1103/PhysRevLett.94.188701.
66. S. Perera, et al., "Structural characteristics of complex supply chain networks," *2017 Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, Sri Lanka, 2017, pp. 135-140, doi: 10.1109/MERCon.2017.7980470.
67. Barabási, AL., Oltvai, Z. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101–113 (2004). https://doi.org/10.1038/nrg1272
68. A.-L. Barabási, Network Science, 1st ed. Cambridge: Cambridge University Press, 2016.
69. Dietz KJ, et al., Hubs and bottlenecks in plant molecular signalling networks. New Phytol. 2010 Dec;188(4):919-38. doi: 10.1111/j.1469-8137.2010.03502.x. Epub 2010 Oct 19. PMID: 20958306.
70. Ermagun, A., et al., Uncertainty in vulnerability of networks under attack. *Sci Rep* **13**, 3179 (2023). https://doi.org/10.1038/s41598-023-29899-w
71. McCabe ERB. Metabolite flux: A dynamic concept for inherited metabolic disorders as complex traits. Mol Genet Metab. 2019 Sep-Oct;128(1-2):14-18. doi: 10.1016/j.ymgme.2019.07.007. Epub 2019 Jul 16. PMID: 31331738.
72. Newman, Mark. "The physics of networks." *Physics today* 61.11 (2008): 33-38. https://doi.org/10.1063/1.3027989
73. Bader G.D. and Hogue,C.W. (2002) Analyzing yeast protein–protein interaction data obtained from different sources. Nat. Biotechnol., 20, 991–997.
74. J. Wang, et al., Low protein diet up-regulate intramuscular lipogenic gene expression and down-regulate lipolytic gene expression in growth finishing pigs. Livestock Science, 2012, Vol 148, Issues 2, Pages 119-128, https://doi.org/10.1016/j.livsci.2012.05.018.
75. Cowley, M., et al., Intra- and inter-individual genetic differences in gene expression. Nat Prec (2008). https://doi.org/10.1038/npre.2008.1799.1
76. Sosa-Acosta, P., et al., (2022). Proteomics of ZIKV infected amniotic fluids of microcephalic fetuses reveals extracellular matrix and immune system dysregulation. Proteomics Clinical Applications, 16, e2100041. https://doi.org/10.1002/prca.202100041
77. Davis, H.E., et al., Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol 21, 133–146 (2023). https://doi.org/10.1038/s41579-022-00846-2
78. Zandi M, et al., The role of SARS-CoV-2 accessory proteins in immune evasion. Biomed Pharmacother. 2022 Dec;156:113889. doi: 10.1016/j.biopha.2022.113889. Epub 2022 Oct 17. PMID: 36265309; PMCID: PMC9574935.
79. Ascenzi D, et al., Stereodynamical Effects by Anisotropic Intermolecular Forces. Front Chem. 2019 May 31;7:390. doi: 10.3389/fchem.2019.00390. PMID: 31214573; PMCID: PMC6554618.
80. Malaney, P., et al. Intrinsic Disorder in PTEN and its Interactome Confers Structural Plasticity and Functional Versatility. Sci Rep 3, 2035 (2013). https://doi.org/10.1038/srep02035
81. Uversky, V. N., et al., Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37, 215–246 (2008).
82. Sharma A, 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. Erratum in: Mol Diagn Ther. 2021 May;25(3):387. PMID: 33475988; PMCID: PMC7847983.

83. Flynn RA, et al., Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions. Cell. 2021 Apr 29;184(9):2394-2411.e16. doi: 10.1016/j.cell.2021.03.012. Epub 2021 Mar 11. PMID: 33743211; PMCID: PMC7951565.
84. Charitou, T., Bryan, K. & Lynn, D.J. Using biological networks to integrate, visualize and analyze genomics data. Genet Sel Evol 48, 27 (2016). https://doi.org/10.1186/s12711-016-0205-1
85. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res. 2013;41:D1228–33.
86. Hong G, et al., Separate enrichment analysis of pathways for up- and downregulated genes. J R Soc Interface. 2013 Dec 18;11(92):20130950. doi: 10.1098/rsif.2013.0950. PMID: 24352673; PMCID: PMC3899863.
87. Wu D, Lim E, et al., (2010). ROAST: rotation gene set tests for complex microarray experiments. Bioinformatics 26, 2176–2182. (10.1093/bioinformatics/btq401)
88. Tavleeva MM, et al., Effects of Antioxidant Gene Overexpression on Stress Resistance and Malignization In Vitro and In Vivo: A Review. Antioxidants (Basel). 2022 Nov 23;11(12):2316. doi: 10.3390/antiox11122316. PMID: 36552527; PMCID: PMC9774954.
89. Sharifi-Rad Mehdi, et al., Lifestyle, Oxidative Stress, and Antioxidants: Back and Forth in the Pathophysiology of Chronic Diseases. Frontiers in Physiology, Review, VOL11, 2020 . DOI=10.3389/fphys.2020.00694. ISSN=1664-042X https://www.frontiersin.org/articles/10.3389/fphys.2020.00694
90. Lu T, et al., Gene regulation and DNA damage in the ageing human brain. Nature. 2004 Jun 24;429(6994):883-91. doi: 10.1038/nature02661. Epub 2004 Jun 9. PMID: 15190254.
91. Jolly MK, et al., EMT and MET: necessary or permissive for metastasis? Mol Oncol. 2017 Jul;11(7):755-769. doi: 10.1002/1878-0261.12083. Epub 2017 Jun 12. PMID: 28548345; PMCID: PMC5496498.
92. Kupsco A, Schlenk D. Oxidative stress, unfolded protein response, and apoptosis in developmental toxicity. Int Rev Cell Mol Biol. 2015;317:1-66. doi: 10.1016/bs.ircmb.2015.02.002. Epub 2015 Mar 11. PMID: 26008783; PMCID: PMC4792257.
93. Priya Wadgaonkar, Fei Chen, Connections between endoplasmic reticulum stress-associated unfolded protein response, mitochondria, and autophagy in arsenic-induced carcinogenesis. Seminars in Cancer Biology, Volume 76, 2021, Pages 258-266, ISSN 1044-579X, https://doi.org/10.1016/j.semcancer.2021.04.004.
94. Bokai Song et al. Intercellular communication within the virus microenvironment affects the susceptibility of cells to secondary viral infections.Sci. Adv.9,eadg3433(2023).DOI:10.1126/sciadv.adg3433
95. 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
96. K. P. Sinaga and M. -S. Yang, "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, pp. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
97. M. P. Rombach, et al., Core-periphery structure in networks, SIAM Journal on Applied Mathematics 74, 167-190 (2014).
98. Laughlin RB, et al., The middle way. Proc. Natl. Acad. Sci. USA. 2000; 97:32–37. [PMC free article] [PubMed] [Google Scholar]
99. Sear RP, Pagonabarraga I, Flaus A. Life at the mesoscale: The self-organised cytoplasm and nucleoplasm. BMC Biophys. 2015;8:4.
100. PJ Thul, et al., A subcellular map of the human proteome. Science, 2017 11 May 2017, Vol 356, Issue 6340 DOI: 10.1126/science.aal3321

101. S Garg, et al., Characterization of methionine dependence in melanoma cells. Molecular Omics, (2023) in press. (bioRxiv preprint doi: https://doi.org/10.1101/2023.04.05.535723.
102. Samad AFA, et al., MicroRNA and Transcription Factor: Key Players in Plant Regulatory Network. Front Plant Sci. 2017 Apr 12;8:565. doi: 10.3389/fpls.2017.00565. PMID: 28446918; PMCID: PMC5388764.
103. Vishnubalaji R, Reciprocal interplays between MicroRNAs and pluripotency transcription factors in dictating stemness features in human cancers. Semin Cancer Biol. 2022 Dec;87:1-16. doi: 10.1016/j.semcancer.2022.10.007. Epub 2022 Oct 29. PMID: 36354097.
104. Raices, M., D'Angelo, M. Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. *Nat Rev Mol Cell Biol* **13**, 687–699 (2012). https://doi.org/10.1038/nrm3461
105. Tang BL, et al., The mammalian homolog of yeast Sec13p is enriched in the intermediate compartment and is essential for protein transport from the endoplasmic reticulum to the Golgi apparatus. Mol Cell Biol. 1997 Jan;17(1):256-66. doi: 10.1128/MCB.17.1.256. PMID: 8972206; PMCID: PMC231750.
106. Betel D, et al., The microRNA.org resource: targets and expression. Nucleic Acids Res. 2008, 36: D149-153.
107. Isakova A, et al., A mouse tissue atlas of small noncoding RNA. Proc Natl Acad Sci U S A. 2020 Oct 13;117(41):25634-25645. doi: 10.1073/pnas.2002277117. Epub 2020 Sep 25. PMID: 32978296; PMCID: PMC7568261.
108. Pérez-Galarza, J. Et al., Immune Response to SARS-CoV-2 Infection in Obesity and T2D: Literature Review. Vaccines 2021, 9, 102. https://doi.org/10.3390/vaccines9020102
109. Claudia Pinacchio, et al., (2022) Analysis of serum microRNAs and rs2910164 GC single-nucleotide polymorphism of miRNA-146a in COVID-19 patients, Journal of Immunoassay and Immunochemistry, 43:4, 347-364, DOI: 10.1080/15321819.2022.2035394
110. Wang, Y., et al. Decreased inhibition of exosomal miRNAs on SARS-CoV-2 replication underlies poor outcomes in elderly people and diabetic patients. Sig Transduct Target Ther 6, 300 (2021). https://doi.org/10.1038/s41392-021-00716-y
111. Hadavi R, et al., Expression of Bioinformatically Candidate miRNAs including, miR-576-5p, miR-501-3p and miR-3143, Targeting PI3K Pathway in Triple-Negative Breast Cancer. Galen Med J. 2019 Nov 10;8:e1646. doi: 10.31661/gmj.v8i0.1646. PMID: 34466540; PMCID: PMC8343935.
112. Su Li, et al., Identification of candidate biomarkers for epithelial ovarian cancer metastasis using microarray data. Oncology Letters, October-2017, Volume 14 Issue 4, Print ISSN: 1792-1074 https://doi.org/10.3892/ol.2017.6707
113. Khokhar M, et al., Novel Molecular Networks and Regulatory MicroRNAs in Type 2 Diabetes Mellitus: Multiomics Integration and Interactomics Study JMIR Bioinform Biotech 2022;3(1):e32437 doi: 10.2196/32437
114. Wang Y, et al., A transcriptional roadmap to the induction of pluripotency in somatic cells. Stem Cell Rev Rep. 2010 Jun;6(2):282-96. doi: 10.1007/s12015-010-9137-2. PMID: 20336394.
115. Aubrey BJ, Strasser A, Kelly GL. Tumor-Suppressor Functions of the TP53 Pathway. Cold Spring Harb Perspect Med. 2016 May 2;6(5):a026062. doi: 10.1101/cshperspect.a026062. PMID: 27141080; PMCID: PMC4852799.
116. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol. 2010 Jan;2(1):a001008. doi: 10.1101/cshperspect.a001008. PMID: 20182602; PMCID: PMC2827900.

117. Lee JD, et al., Differences in syncytia formation by SARS-CoV-2 variants modify host chromatin accessibility and cellular senescence via TP53. bioRxiv [Preprint]. 2023 Sep 1:2023.08.31.555625. doi: 10.1101/2023.08.31.555625. PMID: 37693555; PMCID: PMC10491142.
118. Tiwari R, et al., Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis. J Biomol Struct Dyn. 2022 Aug;40(13):5868-5879. doi: 10.1080/07391102.2021.1874532. Epub 2021 Jan 28. PMID: 33506741; PMCID: PMC7852281.
119. Liu CG, et al., (2008). "MicroRNA expression profiling using microarrays". Nature Protocols. 3 (4): 563–8. doi:10.1038/nprot.2008.14. PMID 18388938. S2CID 2441105.
120. Lagendijk AK, et al., (2012). "Revealing details: whole mount microRNA in situ hybridization protocol for zebrafish embryos and adult tissues". Biology Open. 1 (6): 566–9. doi:10.1242/bio.2012810. PMC 3509442. PMID 23213449.
121. Shingara J, et al., (2005). "An optimized isolation and labeling platform for accurate microRNA expression profiling". RNA. 11 (9): 1461–70. doi:10.1261/rna.2610405. PMC 1370829. PMID 16043497.
122. Chen Z, et al., Interactomes of SARS-CoV-2 and human coronaviruses reveal host factors potentially affecting pathogenesis. EMBO J. 2021 Sep 1;40(17):e107776. doi: 10.15252/embj.2021107776. Epub 2021 Jul 26. PMID: 34232536; PMCID: PMC8447597.
123. Delgado-Chaves FM, et al., Computational Analysis of the Global Effects of Ly6E in the Immune Response to Coronavirus Infection Using Gene Networks. Genes (Basel). 2020 Jul 21;11(7):831. doi: 10.3390/genes11070831. PMID: 32708319; PMCID: PMC7397019.
124. Guzzi PH, et al., Master Regulator Analysis of the SARS-CoV-2/Human Interactome. J Clin Med. 2020 Apr 1;9(4):982. doi: 10.3390/jcm9040982. PMID: 32244779; PMCID: PMC7230814.
125. Zaim S, et al., COVID-19 and Multiorgan Response. Curr Probl Cardiol. 2020 Aug;45(8):100618. doi: 10.1016/j.cpcardiol.2020.100618. Epub 2020 Apr 28. PMID: 32439197; PMCID: PMC7187881.
126. Thakur, V. et al., Multi-Organ Involvement in COVID-19: Beyond Pulmonary Manifestations. J. Clin. Med. 2021, 10, 446. https://doi.org/10.3390/jcm10030446.
127. K. Moravcevic, et al., "Conditional Peripheral Membrane Proteins: Facing up to Limited Specificity" Structure, VOL. 20, ISSUE 1, P15-27, JANUARY 11, (2012). https://doi.org/10.1016/j.str.2011.11.012.
128. Seaton B.A. and Roberts M.F. Peripheral membrane proteins. pp. 355-403. In *Biological Membranes* (Eds. K. Mertz and B.Roux), Birkhauser Boston, 1996.
129. A.M. Whited, A. Johs. The interactions of peripheral membrane proteins with biological membranes. Chemistry and Physics of Lipids Volume 192, November 2015, Pages 51-59. https://doi.org/10.1016/j.chemphyslip.2015.07.015
130. Félix M. Goñi (2002) Non-permanent proteins in membranes: when proteins come as visitors (Review), Molecular Membrane Biology, 19:4, 237-245, DOI: 10.1080/0968768021000035078.
131. Puthenveetil, R.; Christenson, E.T.; Vinogradova, O. New Horizons in Structural Biology of Membrane Proteins: Experimental Evaluation of the Role of Conformational Dynamics and Intrinsic Flexibility. *Membranes* **2022**, *12*, 227. https://doi.org/10.3390/membranes12020227
132. Cho, W. and Stahelin, R.V. (June 2005). "Membrane-protein interactions in cell signaling and membrane trafficking". Annual Review of Biophysics and Biomolecular Structure **34**: 119–151. doi:10.1146/annurev.biophys.33.110502.133337.
133. Robertson RM, et al., A two-helix motif positions the lysophosphatidic acid acyltransferase active site for catalysis within the membrane bilayer. Nat Struct Mol Biol. 2017 Aug;24(8):666-671. doi: 10.1038/nsmb.3436. Epub 2017 Jul 17. PMID: 28714993; PMCID: PMC5616210.

134. *Seelig J (2004). "Thermodynamics of lipid-peptide interactions". Biochim Biophys Acta* ***1666*** *(1-2): 40-50. PMID 15519307.*
135. Allen KN, et al., Monotopic Membrane Proteins Join the Fold. Trends Biochem Sci. 2019 Jan;44(1):7-20. doi: 10.1016/j.tibs.2018.09.013. Epub 2018 Oct 15. PMID: 30337134; PMCID: PMC6309722.
136. Dowran R, et al., Reinfection and reactivation of SARS-CoV-2. Future Virol. 2022 Sep:10.2217/fvl-2021-0212. doi: 10.2217/fvl-2021-0212. Epub 2022 Sep 26. PMID: 36176508; PMCID: PMC9514089.
137. Costanzo M, et al., Deciphering the Relationship between SARS-CoV-2 and Cancer. Int J Mol Sci. 2023 Apr 25;24(9):7803. doi: 10.3390/ijms24097803. PMID: 37175509; PMCID: PMC10178366.
138. Stingi A, Cirillo L. SARS-CoV-2 infection and cancer: Evidence for and against a role of SARS-CoV-2 in cancer onset. Bioessays. 2021 Aug;43(8):e2000289. doi: 10.1002/bies.202000289. Epub 2021 Jun 3. PMID: 34081334; PMCID: PMC8209829.
139. Cardozo C.M., Hainaut P. Viral strategies for circumventing p53: the case of severe acute respiratory syndrome coronavirus. Curr Opin Oncol. 2021;33(2):149–158. PubMed PMID: 33405482. Pubmed Central PMCID: PMC7924916.
140. Gómez-Carballa A, et al., Is SARS-CoV-2 an oncogenic virus? J Infect. 2022 Nov;85(5):573-607. doi: 10.1016/j.jinf.2022.08.005. Epub 2022 Aug 9. PMID: 35961462; PMCID: PMC9361571.
141. Jahankhani K, et al., Possible cancer-causing capacity of COVID-19: Is SARS-CoV-2 an oncogenic agent? Biochimie. 2023 Oct;213:130-138. doi: 10.1016/j.biochi.2023.05.014. Epub 2023 May 23. PMID: 37230238; PMCID: PMC10202899.
142. Carlson CR, et al., **2020**. Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol. Cell* 80:1092–103.e4
143. Cubuk J, et al., **2021**. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* 12:1936.
144. Wu C, Qavi AJ, et al. **2021**. Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain. *iScience* 24:102681
145. Zheng Y, Gao C. Phase Separation: The Robust Modulator of Innate Antiviral Signaling and SARS-CoV-2 Infection. Pathogens. 2023 Feb 3;12(2):243. doi: 10.3390/pathogens12020243. PMID: 36839515; PMCID: PMC9962166.
146. Jack A, et al., **2021**. SARS-CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. *PLOS Biol*. 19:e3001425
147. Lu S, et al. **2021**. The SARS-CoV-2 nucleocapsid phosphoprotein formsmutually exclusive condensates with RNA and the membrane-associatedMprotein. *Nat.Commun.*12:502
148. Risso-Ballester J., et al., A condensate-hardening drug blocks RSV replication in vivo. Nature. 2021;595:596–599. doi: 10.1038/s41586-021-03703-z.
149. Li H., et al., Phase separation in viral infections. Trends Microbiol. 2022;30:1217–1231. doi: 10.1016/j.tim.2022.06.005.
150. Wang B., et al., Liquid-liquid phase separation in human health and diseases. Signal Transduct. Target. Ther. 2021;6:290. doi: 10.1038/s41392-021-00678-1.
151. Li P, et al., 2012. Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483:336–40
152. 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.

153. 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.
154. 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.
155. 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.
156. 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.
157. S. Perera, H. N. Perera and D. Kasthurirathna, "Structural characteristics of complex supply chain networks," *2017 Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, Sri Lanka, 2017, pp. 135-140, doi: 10.1109/MERCon.2017.7980470.
158. A.-L. Barabási, Network Science, 1st ed. Cambridge: Cambridge University Press, 2016.
159. Syakur, M A et al., Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster - IOP Conference Series. Materials Science and Engineering**;** **Bristol** Vol. 336, Fasc. 1, (Apr 2018). DOI:10.1088/1757.
160. Zhou, G., et al., (2019) "NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis" Nucleic Acids Research 47 (W1): W234-W241.
161. Xia J, et al., NetworkAnalyst--integrative approaches for protein-protein interaction network analysis and visual exploration. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W167-74. doi: 10.1093/nar/gku443. Epub 2014 May 26. PMID: 24861621; PMCID: PMC4086107.
162. Zhi-Ping Liu, et al., (2015). RegNetwork: an integrated database of transcriptional and posttranscriptional regulatory networks in human and mouse. Database 2015. doi: 10.1093/database/bav095
163. Gábor Erdős, Zsuzsanna Dosztányi. Analyzing Protein Disorder with IUPred2A. . Current Protocols in Bioinformatics 2020;**70(1)**:e99
164. Drozdetskiy A, et al., JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 2015 Jul 1;43(W1):W389-94. doi: 10.1093/nar/gkv332. Epub 2015 Apr 16. PMID: 25883141; PMCID: PMC4489285.
165. Yuan Ren et al., Force redistribution in clathrin-mediated endocytosis revealed by coiled-coil force sensors. Sci. Adv.9, eadi1535(2023). DOI:10.1126/sciadv.adi1535.
166. M. Skruzny, E. Pohl, S. Gnoth, G. Malengo, V. Sourjik, The protein architecture of the endocytic coat analyzed by FRET microscopy. Mol. Syst. Biol. 16, e9009 (2020).
167. Mahapatra A, et al., The Mechanics and Thermodynamics of Tubule Formation in Biological Membranes. J Membr Biol. 2021 Jun;254(3):273-291. doi: 10.1007/s00232-020-00164-9. Epub 2021 Jan 19. PMID: 33462667; PMCID: PMC8184589.
168. R. C. Aguilar, et al., The yeast epsin Ent1 is recruited to membranes through multiple independent interactions. J. Biol. Chem. 278, 10737–10743 (2003).
169. Å. E. Y. Engqvist-Goldstein, et al., The actin-binding protein Hip1R associates with clathrin during early stages of endocytosis and promotes clathrin assembly in vitro. J. Cell Biol. 154, 1209–1224 (2001).
170. J. D. Wilbur, et al., Actin binding by Hip1 (Huntingtin-interacting protein 1) and Hip1R (Hip1-related protein) is regulated by clathrin light chain. J. Biol. Chem. 283, 32870–32879 (2008).

171. Skruzny M. The endocytic protein machinery as an actin-driven membrane-remodeling machine. Eur J Cell Biol. 2022 Sep-Nov;101(4):151267. doi: 10.1016/j.ejcb.2022.151267. Epub 2022 Aug 5.PMID:35970066.
172. Shin N, et al., SNX9 regulates tubular invagination of the plasma membrane through interaction with actin cytoskeleton and dynamin 2. J Cell Sci. 2008 Apr 15;121(Pt 8):1252-63. doi: 10.1242/jcs.016709. PMID: 18388313.
173. Hsu PP, et al., The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. Science. 2011 Jun 10;332(6035):1317-22. doi: 10.1126/science.1199498. PMID: 21659604; PMCID: PMC3177140.
174. Ekim B, et al., mTOR kinase domain phosphorylation promotes mTORC1 signaling, cell growth, and cell cycle progression. Mol Cell Biol. 2011 Jul;31(14):2787-801. doi: 10.1128/MCB.05437-11. Epub 2011 May 16. PMID: 21576368; PMCID: PMC3133410.
175. Jacinto E, et al., Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive. Nat Cell Biol. 2004 Nov;6(11):1122-8. doi: 10.1038/ncb1183. Epub 2004 Oct 3. PMID: 15467718.
176. Ma G, et al., DEPTOR as a novel prognostic marker inhibits the proliferation via deactivating mTOR signaling pathway in gastric cancer cells. Exp Cell Res. 2023 Jun 1;427(1):113598. doi: 10.1016/j.yexcr.2023.113598. Epub 2023 Apr 11. PMID: 37054772.
177. Information about DEPTOR is also from NCBI – USA (https://www.ncbi.nlm.nih.gov/gene/64798).
178. Wu, W., et al., The SARS-CoV-2 nucleocapsid protein: its role in the viral life cycle, structure and functions, and use as a potential target in the development of vaccines and diagnostics. Virol J 20, 6 (2023). https://doi.org/10.1186/s12985-023-01968-6
179. Kloc, M., et al., Virus interactions with the actin cytoskeleton—what we know and do not know about SARS-CoV-2. Arch Virol 167, 737–749 (2022). https://doi.org/10.1007/s00705-022-05366-1
180. Zhong P., et al., Trasmissione dei virus da cellula a cellula . Curr. Opinion. Virolo. 3 , 44–50 (2013).
181. Dale B. M., et al., Cell-to-cell transfer of HIV-1 via virological synapses leads to endosomal virion maturation that activates viral membrane fusion. Cell Host Microbe 10, 551–562 (2011).
182. Kruglova N, Siniavin A, Gushchin V, Mazurov D. Different Neutralization Sensitivity of SARS-CoV-2 Cell-to-Cell and Cell-Free Modes of Infection to Convalescent Sera. Viruses. 2021 Jun 12;13(6):1133. doi: 10.3390/v13061133. PMID: 34204732; PMCID: PMC8231521.
183. 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.
184. Carlson CR, et al., Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. Mol Cell. 2020;80:1092-103.e4.
185. Zhao M, et al., GCG inhibits SARS-CoV-2 replication by disrupting the liquid phase condensation of its nucleocapsid protein. Nat Commun. 2021;12:2114.
186. Zhao D, et al., Understanding the phase separation characteristics of nucleocapsid protein provides a new therapeutic opportunity against SARS-CoV-2. Protein Cell. 2021;12:734–40.
187. Alizon S, et al., Virulence evolution and the trade-off hypothesis: History, current state of affairs and the future. J Evol Biol. 2009;22(2):245–259. pmid:19196383
188. Miller MR, et al., The evolution of parasites in response to tolerance in their hosts: The good, the bad, and apparent commensalism. Evolution. 2006;60(5):945. pmid:16817535
189. Best A, et al., The coevolutionary implications of host tolerance. Evolution. 2014 May;68(5):1426–1435. pmid:24475902
190. Randolph HE, et al., Holy Immune Tolerance, Batman! Immunity. 2018; 48(6):1074–1076. pmid:29924972

191. Hsieh, Min-Kang, and Jeffery B. Klauda. "Multiscale Molecular Dynamics Simulations of the Homodimer Accessory Protein ORF7b of SARS-CoV-2." *The Journal of Physical Chemistry B* 128.1 (2023): 150-162. Virtual Special Issue "Gregory A. Voth Festschrift".
192. Chen WS et al. Functional independence of the epidermal growth factor receptor from a domain required for ligand-induced internalization and calcium regulation. Cell. 1989 Oct 6;59(1):33-43. doi: 10.1016/0092-8674(89)90867-2. PMID: 2790960.
193. Zhang C. et al. Transmembrane and coiled-coil domain family 1 is a novel protein of the endoplasmic reticulum. PLoS One. 2014 Jan 14;9(1):e85206. doi: 10.1371/journal.pone.0085206. PMID: 24454821; PMCID: PMC3891740.
194. Principe S. et al. In-depth proteomic analyzes of exosomes isolated from expressed prostatic secretions in urine. Proteomics. 2013 May;13(10-11):1667-1671. doi: 10.1002/pmic.201200561. Epub 2013 Apr 23. PMID: 23533145; PMCID: PMC3773505.
195. 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.
196. Appel, E. A. et al. Supramolecular cross-linked networks via host-guest complexation with cucurbituril. *J. Am. Chem. Soc.* **132**, 14251–14260 (2010).
197. Tan, C. S. Y. et al. Distinguishing relaxation dynamics in transiently crosslinked polymeric networks. *Polym. Chem.* **8**, 5336–5343 (2017).
198. Anna V. Ignatenko, et al., Chapter Seven - Nonlinear Chaotic Dynamics of Quantum Systems: Molecules in an Electromagnetic Field, Editor(s): Samantha Jenkins, et al., Advances in Quantum Chemistry, Academic Press, Volume 78, 2019, Pages 149-170, ISSN 0065-3276, ISBN 9780128160848, https://doi.org/10.1016/bs.aiq.2018.06.006. https://www.sciencedirect.com/science/article/pii/S0065327618300303
199. Yu AC, et al., Physical networks from entropy-driven non-covalent interactions. Nat Commun. 2021 Feb 2;12(1):746. doi: 10.1038/s41467-021-21024-7. PMID: 33531475; PMCID: PMC7854746.
200. VanHulle, M M. *Entropy driven artificial neuronal networks and sensorial representation; A proposal*. Journal of Parallel and Distributed Computing; (1989) Vol: 6:2; Journal ID: ISSN 0743-7315
201. Bernaschi, M., et al., The Fitness-Corrected Block Model, or how to create maximum-entropy data-driven spatial social networks. *Sci Rep* **12**, 18206 (2022). https://doi.org/10.1038/s41598-022-22798-6
202. Yongxin Li, et al., Entropy driven circuit as an emerging molecular tool for biological sensing: A review,TrAC Trends in Analytical Chemistry, Volume 134, 2021 , 116142 ISSN 0165-9936, https://doi.org/10.1016/j.trac.2020.116142.
203. Vladimir N. Uversky. Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. Journal of Biological Chemistry (2016) VOLUME 291, ISSUE 13, P6689-6695 DOI:https://doi.org/10.1074/jbc.R115.685859
204. Tanaka, F. & Edwards, S. F. Viscoelastic properties of physically cross-linked networks. 2. Dynamic mechanical moduli. *J. Non-Newton. Fluid Mech.* **43**, 273–288 (1992).
205. Lun XK, Zanotelli VR, et al., Influence of node abundance on signaling network state and dynamics analyzed by mass cytometry. Nat Biotechnol. 2017 Feb;35(2):164-172. doi: 10.1038/nbt.3770. Epub 2017 Jan 16. PMID: 28092656; PMCID: PMC5617104.
206. Chen, Y., Li, Y., Xiong, J. *et al.* Role of *PRKDC* in cancer initiation, progression, and treatment. *Cancer Cell Int* **21**, 563 (2021). https://doi.org/10.1186/s12935-021-02229-8

 doi:10.20944/preprints202403.0473.v1

207. V.N. Uversky, C.J. Oldfield, A.K. Dunker. Intrinsically disordered proteins in human diseases: Introducing the D 2 concept. Annu Rev Biophys, 37 (2008), pp. 215-246, 10.1146/annurev.biophys.37.032807.125924

208. E. Koshland. Application of a Theory of Enzyme Specificity to Protein Synthesis. Proc Natl Acad Sci, 44 (1958), pp. 98-104, 10.1073/pnas.44.2.98