livers

MDPI

Supplementary

Understanding the SARS-CoV-2-human liver interactome using a comprehensive database of the individual virus-host interactions.

By Giovanni Colonna

Network pruning protocol

Table S1. List of original hub genes from the literature, including those shared by multiple articles (142 hub genes)

SERPINE1, IL1RN, THBS1, TNFAIP6, GADD45B, TNFRSF12A, PLA2G7, PTGES, PTX3, GADD45G. MYLK2, FAM83D, STC2, CCDC112, EPHX4, MMP1, ASPM, BUB1B, CDC20, CENPF, CEP55, KIF11, KIF4, NCAPG, NUF2, NUSAP1, PBK, PTTG1, RRM2, TPX2, UBE2C. IL6, IL1B, PTGS2, JUN, FOS, ATF3, SOCS3, CSF3, NFKB2, HBEGF. MMP9, FOS, COL1A2, COL2A1, DKK3, IHH, CYP3A4, PPARGC1A, MMP11, APOD. PDGFRB, MMP14, VWF, CD34, NES, MCAM, CSPG4, MMP1, SPARCL1, MMP10. IL1B, S100A12, FCGR3B, CCR1, S100A8, CCL3, CCL2, CCL4, CLEC4D, LILRA1. ACE, ADAM17, DPP4, TMPRSS2 TNF, AKT1, MAPK14, HIF1A, SP1, IL10, CCL2, CCL5, CXCL10, HAO2, BAAT, SLC27A2, IL6, IL18, IL10, TNF, SOCS1, SOCS3, ICAM1, PTEN, RHOA, GDI2, SUMO1, CASP1, IRAK3, ADRB2, PRF1, GZMB, OASL, CCL5, HSP90AA1, HSPD1, IFNG, MAPK1, RAB5A, TNFRSF1A. ACTB, ATM, CDC42, DHX15, EPRS, GAPDH, HIF1A, HNRNPA1, HRAS, HSP90AB1, HSPA8, IL1B, JUN, POLR2B, PTPRC, RPS27A, SFRS1, SMARCA4, SRC, TNF, UBE2I, VEGFA AKT1, TIMP1, NOTCH, CCNA2, RRM2, TTK, BUB1B, KIF20A, PLK1.

Table S2. Original set stripped of shared genes (126 hub genes)

SERPINE1, IL1RN, THBS1, TNFAIP6, GADD45B, TNFRSF12A, PLA2G7, PTGES, PTX3, GADD45G, MYLK2, FAM83D, STC2, CCDC112, EPHX4, MMP1, ASPM, UB1B, CDC20, CENPF, CEP55, KIF11, KIF4, NCAPG, NUF2, NUSAP1, PBK, PTTG1, RM2, TPX2, UBE2C, IL6, IL1B, PTGS2, JUN, FOS, ATF3, SOCS3, CSF3, NFKB2, BEGF, MMP9, COL1A2, COL2A1, DKK3, IHH, CYP3A4, PPARGC1A, MMP11, APOD, DGFRB, MMP14, VWF, CD34, NES, MCAM, CSPG4, SPARCL1, MMP10, S100A12, CGR3B, CCR1, S100A8, CCL3, CCL2, CCL4, CLEC4D, LILRA1, ACE, ADAM17, PP4, TMPRSS2, AKT1, MAPK14, HIF1A, SP1, IL10, CXCL10, HAO2, BAAT, SLC27A2, IL18, TNF, SOCS1, ICAM1, PTEN, RHOA, GDI2, SUMO1, CASP1, IRAK3, DRB2, PRF1, GZMB, OASL, CCL5, HSP90AA1, HSPD1, IFNG, MAPK1, RAB5A, NFRSF1A, ACTB, ATM, CDC42, DHX15, EPRS, GAPDH, HNRNPA1, HRAS, SP90AB1, HSPA8, POLR2B, PTPRC, RPS27A, SFRS1, SMARCA4, SRC, UBE2I, EGFA, TIMP1, NOTCH, CCNA2, TTK, KIF20A, PLK1.

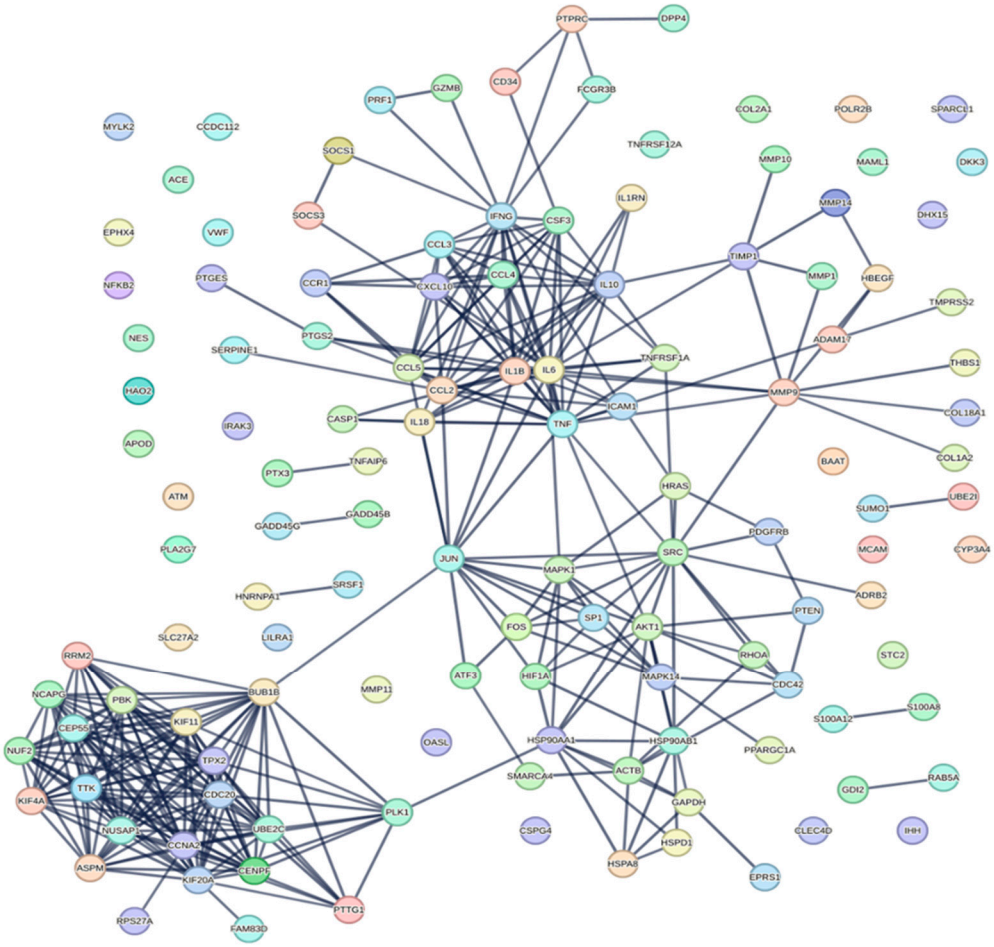

Figure S1. Network of the pruned set of 126 nodes. Number of nodes: 126 number of edges: 343 average node-degree: 5.44 avg. local clustering coefficient: 0.541 expected number of edges: 83 PPI enrichment p-value: $< 1\times10^{-16}$. Data source: all 6 channels; score 0.9.

Figure S1 is followed by Figure S2 where the nodes of the network of Figure S1 will be enriched by the human proteome to extract functional/physical relationships and reduce the number of unconnected nodes.

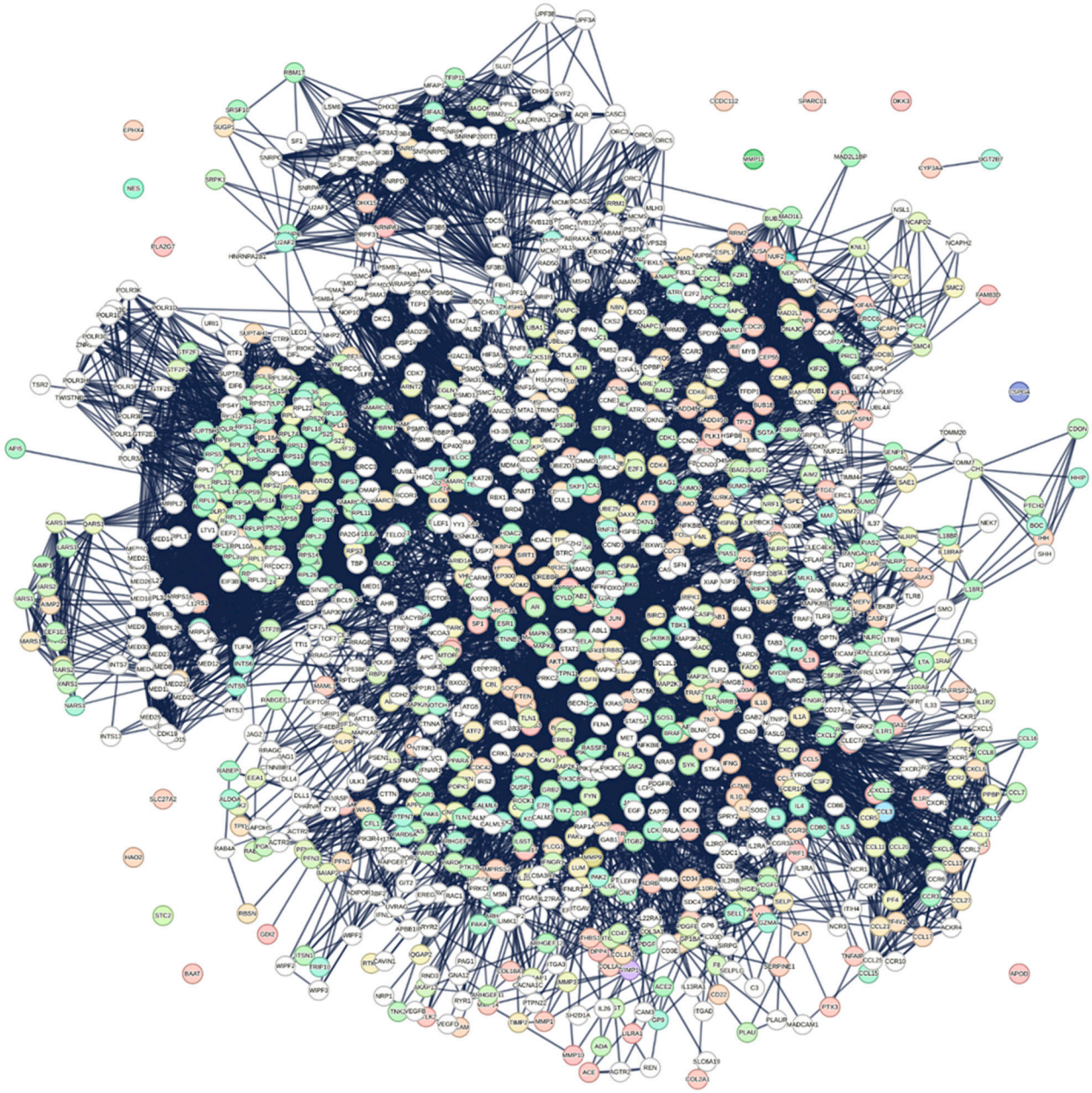

Figure S2. Enriched network of the 126 original hub genes. Number of nodes: 1126; number of edges: 13,483; average node degree: 23.9; avg. local clustering coefficient: 0.614; expected number of edges: 8923; PPI enrichment p-value: < 1×10^{-16}. Data source: all 6 channels; score: 0.9. Enrichment: + 500 first order nodes (direct) + 500 second order nodes (indirect).

Despite the notable physical/functional enrichment, 15 nodes, of which 13 are original parents, are not connected. They were eliminated because any topological analysis gives reliable results only if the analyzed network component is unique. Unconnected nodes (CYP3A4, APOD, BAAT, CCDC112, CSPG4, DKK3, EPHX4, HAO2, MMP11, NES, PLA2G7, LC27A2, SPARCL1, STC2) were manually eliminated from the network via a specific STRING function. We have entered STRING with the remaining 111 original hub-proteins, by adding an enrichment of 1000 proteins to show a network around this input (this occurs by default). The final compact interactome is shown in **Figure 1 of the article**.

Additional data on the interactome in Figure 1.

Table S3. Comprehensive set of enriched functions of the interactome in Figure 1 of the article.

Biological Process (Gene Ontology)	2344 GO-terms significantly enriched
Molecular Function (Gene Ontology)	253 GO-terms significantly enriched
Cellular Component (Gene Ontology)	279 GO-terms significantly enriched
Reference publications (PubMed)	**10,000 publications significantly enriched**
Local network cluster (STRING)	307 clusters significantly enriched
KEGG Pathways	195 pathways significantly enriched
Reactome Pathways	960 pathways significantly enriched
Wiki-Pathways	432 pathways significantly enriched
Disease-gene associations (DISEASES)	273 diseases significantly enriched
Tissue expression (TISSUES)	325 tissues significantly enriched
Subcellular localization (COMPARTMENTS)	340 compartments significantly enriched
Human Phenotype (Monarch)	1406 phenotypes significantly enriched
Annotated Keywords (UniProt)	105 keywords significantly enriched
Protein Domains and Features (InterPro)	75 domains significantly enriched
Protein Domains (SMART)	19 domains significantly enriched
All enriched terms (without PubMed)	7313 enriched terms in 14 categories

Note: It is important to consider that STRING has reviewed ten thousand scientific articles (in red) containing information and data on all the nodes present, in order to be able to calculate the functional/structural relationships existing in the metabolic context defined by the interactome.

Two observations: 1) the interactome of figure 1 covers a large set of metabolic features (7,313). 2) The knowledge base on which the calculations that generated the interactome in figure 1 were carried out derives from over 10,000 scientific articles specific to the context in question. All data and scientific information were extracted from these articles and used for calculations by STRING. The composition of the sources that contribute to the calculation of each single interaction with a score of 0.900 is reported in EXCEL FILE S3.

Brief note on the rationale for calculating of the expected statistical number of interactions and on the network bias.

Proteins in a network should be no more likely to interact with each other than random proteins. When enrichment is implemented, the networks present enrichment distortions because of the interaction of a greater degree for proteins that are studied more often than others. Structural characterization of the human interactome has lagged, and less than 5% of hundreds of thousands of human protein interactions have been experimentally characterized in terms of structure/function [120].

Molecular machines that assemble through protein-protein interactions govern cellular functions. Protein interactions range from transient functional interactions, which regulate enzymatic activity, to permanent interactions in molecular machines. Proteins are, therefore, the fundamental cellular effectors in determining almost all cellular processes. These processes act in a coordinated manner, and the coordination of the many and diverse processes arises from the interaction between proteins and other biomolecules. The quantitative characterization of protein-protein interactions (PPI) is therefore fundamental to understand which groups of proteins form functional units induced by the virus [121]. Although the use of neural networks has demonstrated the ability to predict the structures of single proteins [122,123] and protein complexes [124–

126] successfully, we could not predict the structures of most dimers [124] in Saccharomyces cerevisiae [126] correctly. Therefore, the application of neural networks for large-scale prediction of patterns of human complex structures is yet to come and has not yet been tested.

It is customary for archive curators to collapse the functional characteristics of the proteoforms onto the gene or the native protein it encodes. STRING is no exception and warns of this. Proteoforms are molecules different from the native protein because they undergo chemical and functional modifications, often very numerous. They carry out their function at different cellular times and places. Collapsing their characteristics onto the "native" node means altering and distorting the resulting interactome because that node, being highly functional, will express more relationships than necessary and, therefore, will have an abnormally higher degree. As a protein's functional multiplicity is studied more, its degree in the networks will increase arbitrarily. These biases distort the quantitative results of the networks and their topology.

The resulting PPI networks are valuable for identifying proteins that are relevant to a biological process (49, 50), and their proteins are no more likely to interact with each other than proteins from a similar random network. In contrast, after enrichment, virus-induced human genes represent a more biologically coherent set, encode proteins that interact with each other, and can predict new virus-induced genes and interactions. Therefore, STRING considers a set of proteins physically and biologically cohesive by quantifying that they have more interactions with each other than proteins in random networks. All this is essential to get reliable network medicine, but above all, to get reliable and meaningful calculations of topological parameters.

Clustering

Clusters calculated by STRING on the interactome of Figure 1.

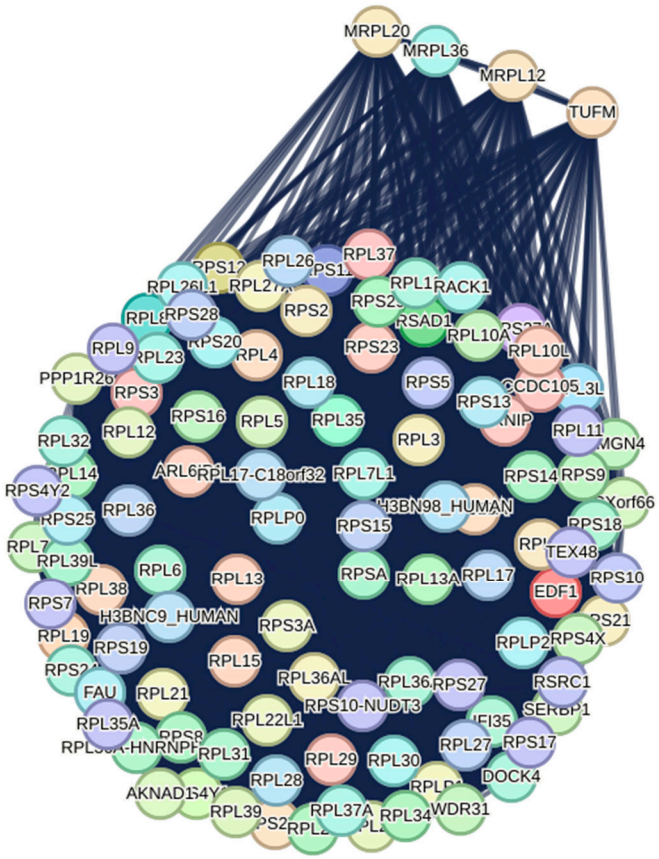

Figure S3. CLUSTER CL:143. Viral mRNA Translation, and Sec61 translocon complex (FDR: 4.07×10^{-223}) Number of nodes: 108; number of edges: 5537; average node degree: 103; avg. local clustering coefficient: 0.978; expected number of edges: 940; PPI enrichment p-value: $< 1\times10^{-16}$.

Table S4. Functional characteristics of Cluster-CL:143

CLUSTER CL:143	Functional activities	Strength	Statistics
GO:0002181	Cytoplasmic translation	2.05	FDR: 3.56×10^{-128}
GO:0022625	Cytosolic large ribosomal subunit	2.2	FDR:2.52×10^{-84}
GO:0022626	Cytosolic ribosome	2.18	FDR: 2.53×10^{-151}
GO:0022627	Cytosolic small ribosomal subunit	2.15	FDR: 1.20×10^{-55}
HSA-156902	Peptide chain elongation	2.23	FDR: 1.05×10^{-149}
HSA-192823	Viral mRNA Translation	2.23	FDR: 1.05×10^{-149}
HSA-72764	Eukaryotic Translation Termination	2.21	FDR: 4.47×10^{-149}
GOCC:0022626	Cytosolic ribosome	2.2	FDR: 1.96×10^{-139}
GOCC:0005840	Ribosome	1.88	FDR: 1.55×10^{-126}
GOCC:0044391	Ribosomal subunit	1.9	FDR: 2.06×10^{-103}
GOCC:1990904	Ribonucleoprotein complex	1.35	FDR: 2.07×10^{-75}

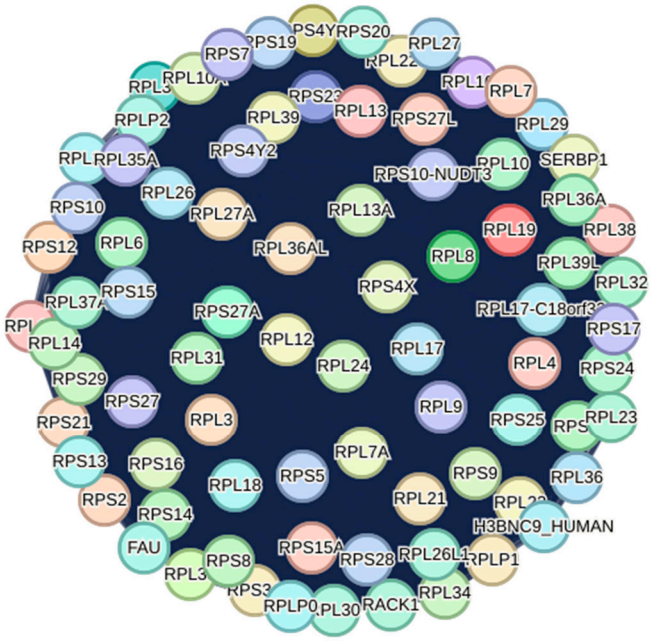

Figure S4. CLUSTER CL:152. Viral mRNA Translation (FDR: 1.16e-176).
Number of nodes: 81; number of edges: 3240; average node degree: 80; avg. local clustering coefficient: 1; expected number of edges: 566; PPI enrichment p-value: $< 1\times10^{-16}$.

Table S5. Functional characteristics of Cluster-CL:152

CLUSTER CL:152	Functional activities	Strength	Statistics
GO:0002181	Cytoplasmic translation	2.15	FDR: 3.19×10^{-133}
GO:0003735	Structural constituent of ribosome	2.05	FDR: 1.84×10^{-146}
GO:0022626	Cytosolic ribosome	2.26	FDR: 6.73×10^{-154}
GO:0022627	Cytosolic small ribosomal subunit	2.23	FDR: 3.81×10^{-54}
HAS-03010	Ribosome	2.15	FDR: 4.04×10^{-148}
HSA-156902	Peptide chain elongation	2.32	FDR: 5.44×10^{-154}
HSA-72764	Eukaryotic Translation Termination	2.3	FDR: 2.03×10^{-153}
GOCC:0022626	Cytosolic ribosome	2.29	FDR: 1.24×10^{-141}
GOCC:0022625	Cytosolic large ribosomal subunit	2.28	FDR: 3.71×10^{-53}
GOCC:0022627	Cytosolic small ribosomal subunit	2.28	FDR: 2.74×10^{-49}

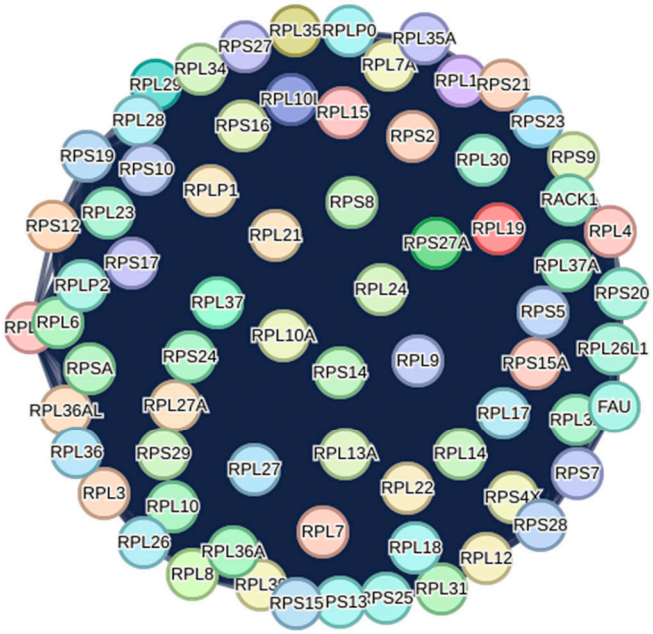

Figure S5. CLUSTER CL:159. Viral mRNA Translation (FDR: 1.58e-1.58)
Number of nodes: 71; number of edges: 2485; average node degree: 70; avg. local clustering coefficient: 1; expected number of edges: 470; PPI enrichment p-value: $< 1\times10^{-16}$.

Table S6. Functional characteristics of Cluster-159

CLUSTER CL.159	Functional activities	Strength	Statistics
GO:0006412	Translation	1.7	FDR: 3.25×10^{-115}
GO:0003735	Structural constituent of ribosome	2.06	FDR: 2.97×10^{-134}
GO:0005925	Focal adhesion	1.38	FDR: 4.17×10^{-39}
KW-0689	Ribosomal protein	2.05	FDR: 3.12×10^{-134}

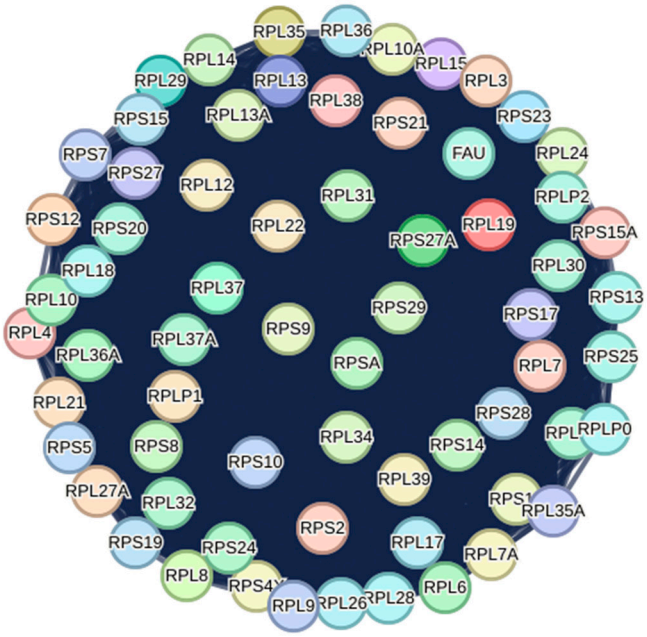

Figure S6. CLUSTER CL:162. Cytoplasmic ribosomal proteins (FDR: 1.67×10^{-98})
Number of nodes: 66; number of edges: 2145; average node degree: 65; avg. local clustering coefficient: 1; expected number of edges: 412; PPI enrichment p-value: $< 1\times10^{-16}$.

Table S7. Functional characteristics of Cluster-CL:162

CLUSTER CL:162	Functional activities	Strength	Statistics
GO:0042254	Ribosome biogenesis	1.43	FDR: 2.44×10^{-28}
GO:0042274	Ribosomal small subunit biogenesis	1.79	FDR: 7.34×10^{-21}
GO:0003735	Structural constituent of ribosome	2.07	FDR: 2.03×10^{-128}
GO:0003723	RNA binding	1.06	FDR: 2.51×10^{-62}
GO:0044391	Ribosomal subunit	2.02	FDR: 1.22×10^{-126}