Molecular Therapy
Nucleic Acids
Original Article

Unveiling the role of PUS7-mediated pseudouridylation in host protein interactions specific for the SARS-CoV-2 RNA genome

Roberto Giambruno,[1,2,13] Elsa Zacco,[3,13] Camilla Ugolini,[1,11,13] Andrea Vandelli,[3,4,5] Logan Mulroney,[1,6,7] Manfredi D'Onghia,[1] Bianca Giuliani,[1] Elena Criscuolo,[8] Matteo Castelli,[8] Nicola Clementi,[8,9] Massimo Clementi,[8,9] Nicasio Mancini,[8,9] Tiziana Bonaldi,[10,11] Stefano Gustincich,[3] Tommaso Leonardi,[1] Gian Gaetano Tartaglia,[3,12,13] and Francesco Nicassio[1,13]

[1]Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia, 20139 Milano, Italy; [2]Institute of Biomedical Technologies, National Research Council, 20090 Segrate, Italy; [3]Central RNA and RNA Systems Biology Labs, Centre for Human Technologies (CHT), Istituto Italiano di Tecnologia (IIT), 16152 Genova, Italy; [4]Department of Biochemistry and Molecular Biology, Universitat Autònoma de Barcelona, Bellaterra, Barcelona 08193, Spain; [5]Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain; [6]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridgeshire CB10 1SD, UK; [7]Epigenetics and Neurobiology Unit, European Molecular Biology Laboratory (EMBL), Monterotondo, RM 00015, Italy; [8]Laboratory of Microbiology and Virology, Vita-Salute San Raffaele University, 20132 Milan, Italy; [9]Laboratory of Medical Microbiology and Virology, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy; [10]Department of Experimental Oncology, European Institute of Oncology IRCCS, 20139 Milano, Italy; [11]Department of Oncology and Hematology-Oncology, University of Milan, 20122 Milano, Italy; [12]Catalan Institution for Research and Advanced Studies, ICREA, 08010 Barcelona, Spain

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a positive single-stranded RNA virus, engages in complex interactions with host cell proteins throughout its life cycle. While these interactions enable the host to recognize and inhibit viral replication, they also facilitate essential viral processes such as transcription, translation, and replication. Many aspects of these virus-host interactions remain poorly understood. Here, we employed the *cat*RAPID algorithm and utilized the RNA-protein interaction detection coupled with mass spectrometry technology to predict and validate the host proteins that specifically bind to the highly structured 5′ and 3′ terminal regions of the SARS-CoV-2 RNA. Among the interactions identified, we prioritized pseudouridine synthase PUS7, which binds to both ends of the viral RNA. Using nanopore direct RNA sequencing, we discovered that the viral RNA undergoes extensive post-transcriptional modifications. Modified consensus regions for PUS7 were identified at both terminal regions of the SARS-CoV-2 RNA, including one in the viral transcription regulatory sequence leader. Collectively, our findings offer insights into host protein interactions with the SARS-CoV-2 UTRs and highlight the likely significance of pseudouridine synthases and other post-transcriptional modifications in the viral life cycle. This new knowledge enhances our understanding of virus-host dynamics and could inform the development of targeted therapeutic strategies.

INTRODUCTION

The outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the human population has dramatically affected life expectancy worldwide. SARS-CoV-2 infection causes the coronavirus disease 2019 (COVID-19), a potentially fatal condition, especially for the elderly population and for individuals with underlying health issues. To date, vaccination is the most effective strategy to significantly decrease the incidence of SARS-CoV-2 infection and the onset of severe COVID-19. However, new and fast-spreading SARS-CoV-2 variants might be able to escape the antibody response of vaccinated people.[1]

SARS-CoV-2 is a positive-sense single-stranded RNA virus of the *Coronoviridae* family.[2,3] Its genome is composed of approximately 30,000 nucleotides and contains 14 open reading frames (ORFs) encoding 16 non-structural proteins (nsp 1–16), 12 accessory proteins (ORF3a, ORF3b, ORF3c, ORF3d, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF9c, ORF9d, and ORF10) and 4 structural proteins: spike, envelope, membrane, and nucleocapsid.[4–6] Structural and accessory proteins are encoded in subgenomic RNAs (sgRNAs) that bear a common 5′ leader sequence fused to different body sequences upstream to the encoded ORF.[7] During its life cycle, SARS-CoV-2

Received 8 September 2023; accepted 5 October 2023;
https://doi.org/10.1016/j.omtn.2023.102052.

[13]These authors contributed equally

Correspondence: Roberto Giambruno, Institute of Biomedical Technologies, National Research Council, 20090 Segrate, Italy.
E-mail: roberto.giambruno@cnr.it
Correspondence: Gian Gaetano Tartaglia, RNA Systems Biology Lab, Istituto Italiano di Tecnologia (IIT), 16152 Genova, Italy.
E-mail: gian.tartaglia@iit.it
Correspondence: Francesco Nicassio, Center for Genomic Science of IIT@SEMM, Fondazione Istituto Italiano di Tecnologia, 20139 Milano, Italy.
E-mail: francesco.nicassio@iit.it

RNAs interact with several host proteins specifically required for the translation of viral proteins, the transcription of the sgRNAs, the replication of the genome, and the generation of new viral particles.[8] At the same time, host proteins recognize the presence of the SARS-CoV-2 RNA in the cytoplasm, activating the innate immune response through the interferon signaling pathways.[9–11] Although the interaction between viral RNAs and host proteins has been recognized as a key molecular process for virulence, a complete and clear understanding of these interactions and their underlying mechanisms has not yet been achieved. Hence, our understanding on how viral RNA exploits cellular mechanisms for its own advantage remains limited.

Several laboratories have independently mapped the SARS-CoV-2 RNA-host protein interactome in mammalian cell lines that are permissive to SARS-CoV-2 infection.[12–15] In these studies, the interactome was analyzed using an RNA-centric approach to specifically purify the viral RNA at 8–24 h after infection, a time frame in which the virus is actively replicating and produces viral proteins that hijack the host innate immune response.[16,17] Many host RNA-binding proteins (RBPs) have been reported to interact with the SARS-CoV-2 RNA. These RBPs are involved in multiple biological processes such as RNA splicing, RNA metabolism, nonsense-mediated decay, translation and viral processing. However, the interactions between host proteins and viral RNAs at earlier stages of the viral life cycle and the RNA regions specifically bound by the host proteins remain largely undetected. With this work, we aim to fill these gaps by expressing defined SARS-CoV-2 RNA fragments in human living cells and analyzing those interactions with the host proteins that might occur in the absence of viral proteins. In particular, we focused on the highly structured regions containing the 5′ and the 3′ ends of SARS-CoV-2, which we previously predicted to be interaction hot spots for the host proteins, as observed also for other coronaviruses.[18]

To thoroughly investigate the interactome of these key SARS-CoV-2 RNA regions, we used an approach based on the combination of the *in house* algorithm *cat*RAPID and a well-established proximity-ligation approach,[19] RNA-protein interaction detection-mass spectrometry (RaPID-MS).[20] In addition to the host proteins already reported, we unveiled other RBPs that may have key functions in the early phases of viral infection by interacting with SARS-CoV-2 genomic RNA. Interestingly, we found also the enzyme pseudouridine synthase 7 (PUS7), an RNA modifier that catalyzes the pseudouridylation of RNA transcripts, highlighting a cellular mechanism in charge of introducing post-transcriptional modification on viral RNA. Nanopore direct RNA sequencing of SARS-CoV-2 RNA from infected mammalian cells confirmed the presence of many RNA modifications, including pseudouridylation sites on the viral RNAs. In particular, we found PUS7 consensus sequences heavily modified in the SARS-CoV-2 sgRNAs, with one present in the stem loop 2 of the SARS-CoV-2 5′ UTR, within the transcription regulatory sequence-leader (TRS-L). Biochemical and biophysical assays support a role for this unique modification by influencing both RNA structure and interaction propensity. In conclusion, we propose that the interaction and the activity of cellular pseudouridine synthases may represent a new mechanism able to influence SARS-CoV-2 life cycle.

RESULTS

Prediction of SARS-CoV-2 regions with the greatest propensity to interact with host proteins

We previously reported that RNA regions with high structural content tend to tightly interact with a large number of proteins,[21,22] which suggests that structured regions of the SARS-CoV-2 RNA genome might play a crucial role in its regulation.[18] In line with this finding, we and others have observed that the first and last 1.5 kb of the viral genome recruit specific factors required for viral translation and replication and that they act as targets for the intracellular host defense response.[18,23] Despite previous attempts, a comprehensive and reproducible map of RNA-host protein interactions is missing.[24] We decided to accurately map the regions with the greatest propensity to interact with host proteins and, accordingly, divided the initial as well as terminal 1.5-kb regions of the viral genome into 500-nucleotide-long RNA fragments. Each fragment overlaps the following one by 250 nucleotides. A total of five fragments for the 5′ region (numbered 1–5) and five for the 3′ region (numbered 6–10) was identified (Table S1; Figure 1A). Through the *cat*RAPID algorithm,[25] which estimates the binding affinity of protein-RNA pairs,[26,27] we predicted the interactions of each fragment with human RBPs (>2,000 entries) (Table S2). In this analysis, we did not consider the cell-specificity pattern of expression of the RBPs, as the calculations serve to provide an estimate of the protein-binding ability of RNA fragments. We identified fragments 1 and 2 and fragments 9 and 10 as those harboring the highest scores (Figure 1B). Considering that the 5′ and 3′ UTRs (fragments 1 and 10) have the strongest scores, we reasoned that these regions might also be the most functionally relevant.[28] Altogether, these results provide important indications on the location of structural hot spots potentially relevant for the functional interactions between the virus and the host.

Identification of the human interactome for the 5′ and 3′ ends of the SARS-CoV-2 genome

To analyze the RNA interaction with the host proteins experimentally, we exploited the RaPID-MS approach[20] using the 10 above-described fragments generated by *cat*RAPID (Figures 1A and S1). The fragments were cloned within BoxB sequences (materials and methods), used as tags for the RaPID assay, and were expressed into HEK293T cells as host cell line (Figure 1C). As a negative control, we generated a scrambled sequence of 500 nt of similar GC content compared to the other fragments (named “Scramble” henceforth). The BoxB-tagged EDEN15 RNA was selected as positive control.[20] We verified that (i) each plasmid expressing an RNA fragment was co-expressed with a plasmid coding for the biotin protein ligase λN-HA-BASU by FACS analysis (Figure S2) and (ii) λN-HA-BASU was expressed correctly and able to biotinylate protein substrates in the presence of exogenous biotin by western blot (WB) analysis (Figure S3A). The resulting biotinylated proteins were purified under denaturing conditions using streptavidin beads, as confirmed by WB analysis (Figure S3B), and analyzed by liquid chromatography-tandem MS (LC-MS/MS). In total, 3 independent biological replicates

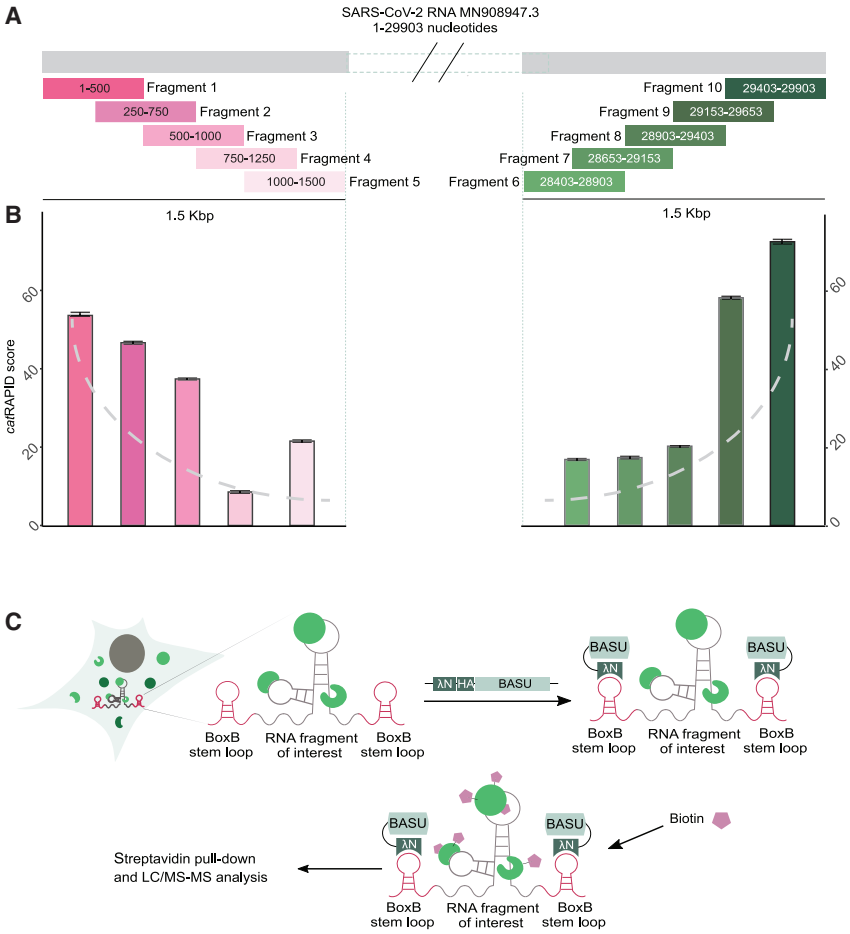


Figure 1. Experimental approach to investigate SARS-CoV-2 interactome

(A) Schematic representation of the SARS-CoV-2 RNA fragments selected to be studied with the RaPID-MS strategy. The scheme shows the fragments names, positions within the SARS-CoV-2 genome and degree of overlap between them. (B) *cat*RAPID predictions of the 10 selected RNA fragments with the catalog of human RBPs. Fragments belonging to the first and last 1.5 Kbp of SARS-CoV-2 genome are colored in different shades of pink and green, respectively. Error bars for each fragment correspond to the average value ± SE. The gray dashed line indicates the trend of the *cat*RAPID score in the chart. (C) Description of the technique RaPID-MS, in which the RNA fragment of interest is expressed in cells flanked by BoxB stem loops. BoxB is specifically recognized by the co-transfected λN peptide fused to the biotin ligase BASU. Upon biotin addition to the growth medium, BASU biotinylates the host protein interactors attracted by the RNA of interest.

of RaPID-MS were performed using the 10 SARS-CoV-2 fragments, as well as the Scramble and the EDEN15 control RNAs. We identified a total of 1,296 proteins interacting with our selected fragments, a resource reported in Table S3. A comparison between our dataset and the ones generated by previous studies indicates an overlap of approximately 40% (Table S3), demonstrating good agreement with the literature.[29,30]

As we are interested in finding host proteins that specifically bind to SARS-CoV-2 regions rather than other RNAs, we compared the list of interactors obtained with each individual fragment with those obtained with our negative control Scramble (Figures S4A and S4B; Table S3). We removed RBPs that were found as interacting to the same extent of Scramble (Table S3), including PTBP1, G3BP1 and SYNCRIP, which were previously reported to specifically interact with the SARS-CoV-2 RNA.[12–15] We were able to verify the specificity of our approach, as we identified CELF1 as the interactor of our positive control EDEN15 (Figure S4B), as expected.[31] In total, we isolated 73 proteins significantly and specifically interacting with SARS-CoV-2 RNA fragments, hereafter called the "RaPID-MS dataset" (Figure 2). In the RaPID-MS dataset, we have 10 proteins, previously reported as interactors of SARS-CoV-2 RNA, namely FAM120A, HAT1, LSG1, SHMT1, and SYNE2 and the ribosomal proteins RPL14, RPL18A, RPL24, RPL35, and RPS6. The remaining proteins were identified for the first time (Figures S5A and S5B). The greatest number of interacting proteins were in fragments 1, 4, 7 and 10. Conversely, fragment 2 displayed only two specific interactions (Figure 2A).

To balance the possible bias that could be introduced by the RaPID technique, we considered protein length and abundance to compute the median enrichment value of each significant interactor (Figure 2B). As for other studies of SARS-CoV-2 RNA interactome,[24] we retrieved ribosomal proteins (N = 7) among the most highly enriched interactors (Figure 2B), explained by the fact that some of the fragments contain ORFs. We expected the RaPID-MS dataset to be enriched for direct RNA-protein interactions. Indeed, 73% (53/73) interactors are annotated as RBPs according to the RBPome database (https://rbpbase.shiny.embl.de/) (green circles in Figure 2C), while the remaining are members of the protein complexes biotinylated in proximity of the target. We noted that, while some proteins (N = 26) were shared among two or more fragments, most interactions (N = 47) are specific for just one RNA fragment (Figure 2C). In some cases, the proteins interact with the overlapping regions included within the fragments, reinforcing our conclusions. Examples are WNK3 and AP4M1, which were found in association with the neighboring fragments 4 and 5 and fragments 7 and 8, respectively. Interestingly, three proteins (CEP350, RGPD1, and GPKOW) were exclusively interacting with the two terminal fragments (1 and 10), suggesting they might bind 5′ and 3′ ends either independently or by participating in viral RNA circularization, as recently described.[32] Gene set enrichment analysis for Gene Ontology (GO) processes highlighted that the retrieved interactors

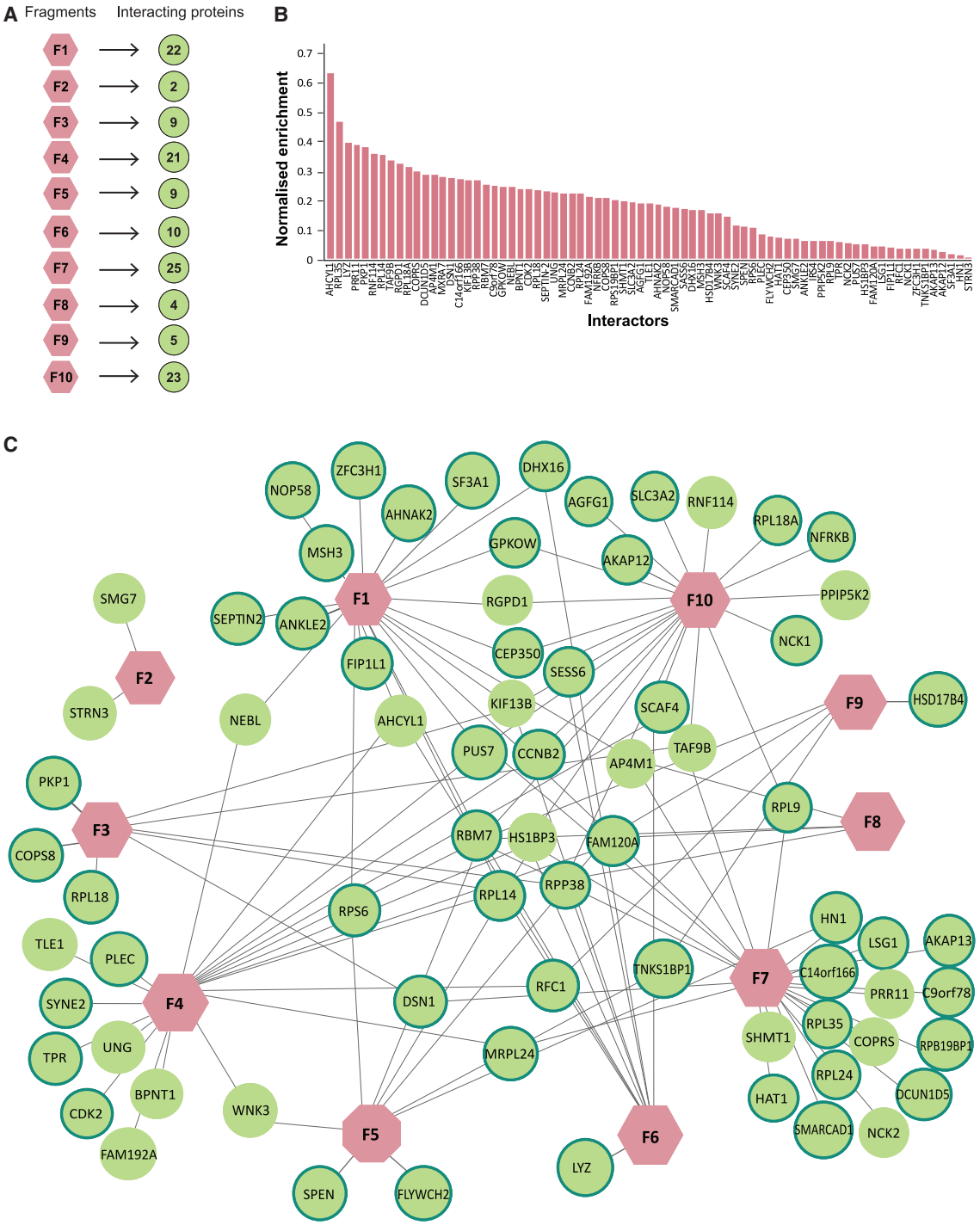
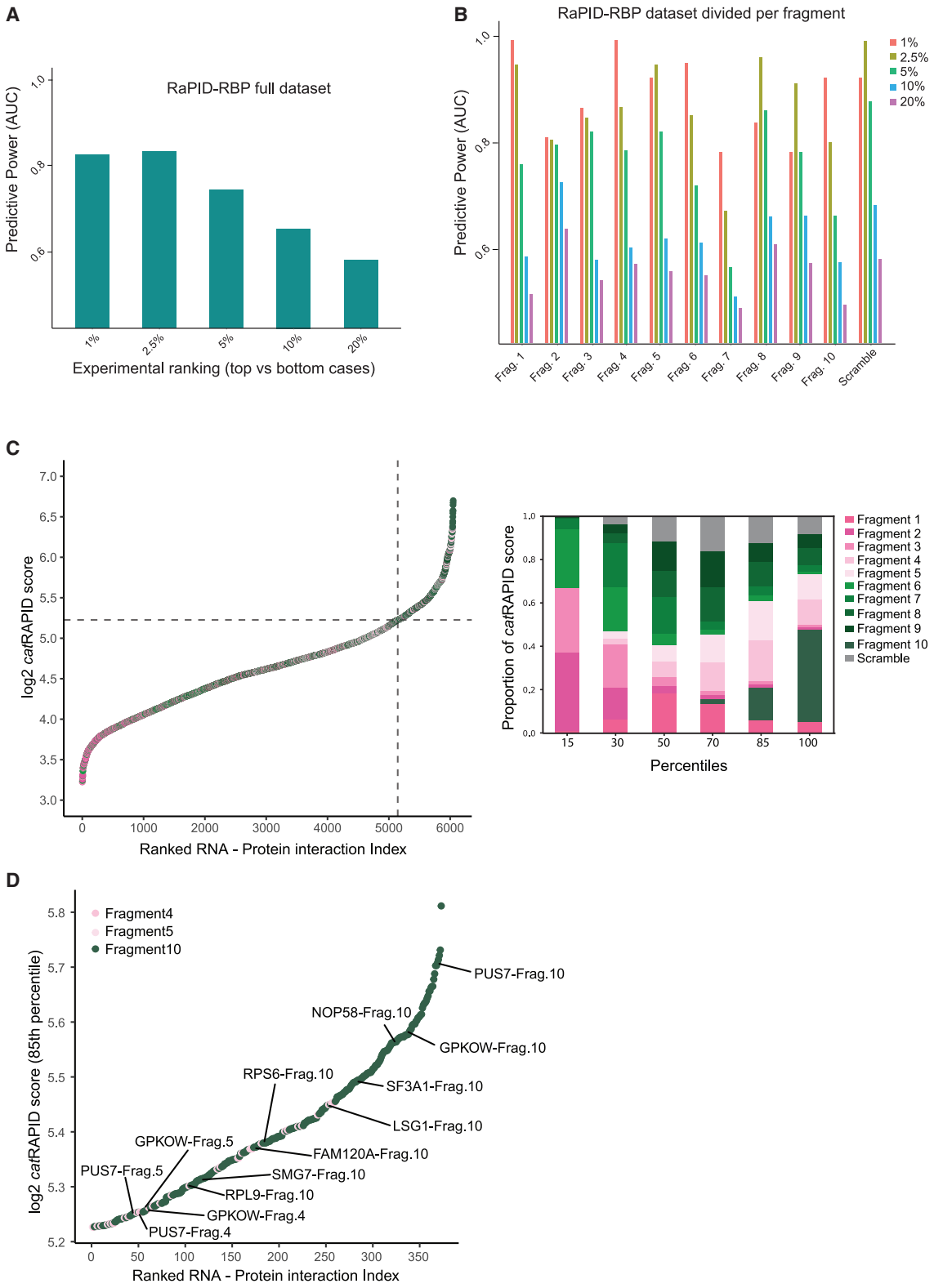


Figure 2. Defining SARS-CoV-2 interaction network with the human proteome

(A) Number of specific interactors identified for each SARS-CoV-2 RNA fragment (labeled with the letter "F" follow by a number). (B) Enrichment distribution of the specific interactors for all considered fragments compared with the control "Scramble." Data are normalized for protein length and abundance, as specified in the materials and methods. (C) The network of proteins identified by RaPID-MS. Only proteins significantly enriched over the control "Scramble" are displayed. Data are derived from the analysis of three independent biological replicates. In pink are displayed the RNA fragments, while in green the retrieved interactors. Proteins circled in green are RBPs, according to the RBPome database (https://rbpbase.shiny.embl.de/).



(legend on next page)

belonged to cellular pathways involved in viral transcription, nonsense-mediated mRNA decay, rRNA processing, translational initiation, mRNA splicing, and mRNA export from the nucleus (Figure S5C).

Correlation between computationally predicted and experimentally validated RNA-host protein interactors

To properly compare predicted and experimental interactions, the median enrichment value of each experimentally identified interactor was normalized by protein length and abundance (Table S4). We restricted the computational prediction analysis to human RBPs reported in the *cat*RAPID library,[25] here named RaPID-RBP (Table S4), and assessed to what extent the predicted binding propensities are in agreement with the experimental results obtained in the RAPID-MS dataset. In particular, we verified whether the strongest positive (i.e., interacting) and strongest negative (i.e., non-interacting) predicted protein-RNA pairs could be identified in the list of experimentally validated interactions. We evaluated the performance of our prediction using the area under the receiver operating characteristic curve (AUC). Overall, *cat*RAPID reached an AUC of more than 0.80. The AUC increased from 0.58 to 0.83 when applying the predictions to the experimental scores from the top 20% (i.e., the 20% strongest positives vs. the 20% strongest negatives) with the top 2.5% (i.e., the 2.5% strongest positives vs. the 2.5% strongest negatives) (Figure 3A; Table S4). When considering each fragment separately, the prediction performances increased even further for the terminal regions of the viral genome (fragments 1 and 10) and for some internal overlapping fragments (fragments 4–6), reaching an AUC of 0.95 for the top 1% ranked experimental cases (Figure 3B; Table S4). Overall, this analysis suggests that predicted and the experimental data are well correlated.

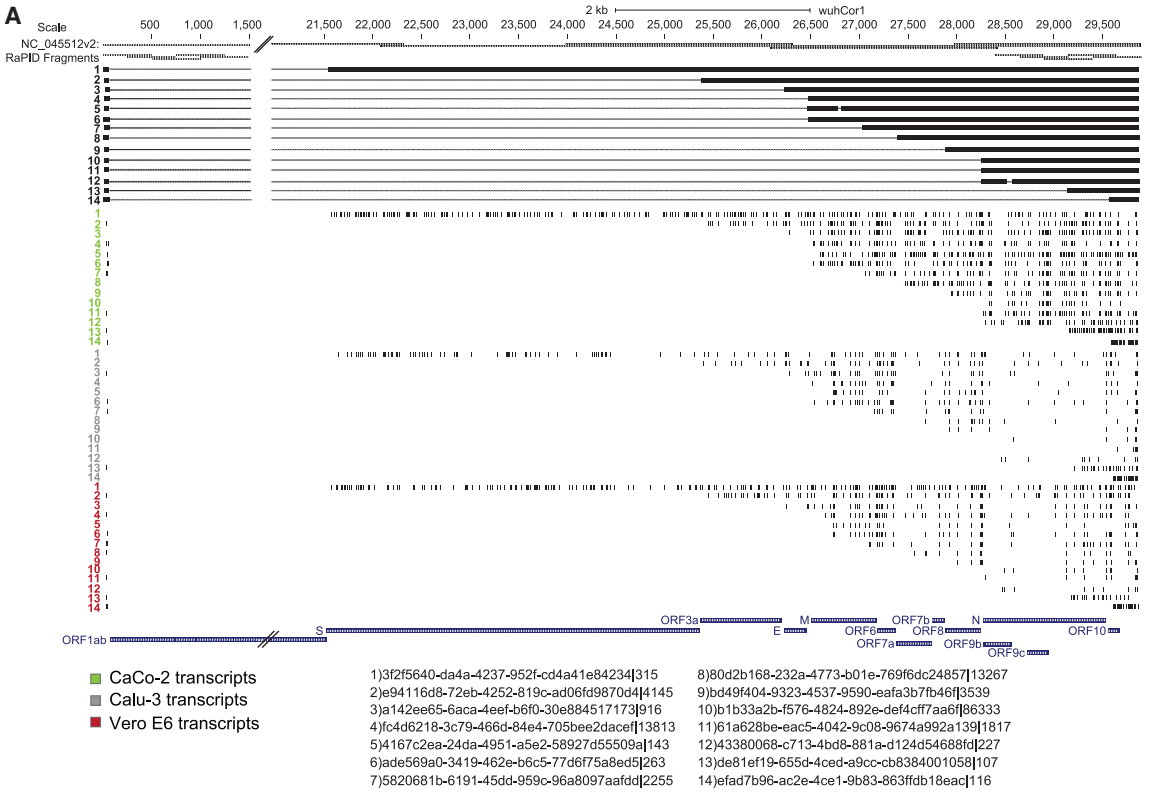
Prompted by this observation, we exploited the *cat*RAPID score to produce a rank of the RBP interactome of SARS-CoV-2 RNA identified by RAPID-MS. The $\log_2$ *cat*RAPID score for all the possible viral RNA-host protein interactions was between 3.22 and 6.70, with a median value of 4.62 (Figure 3C; Table S4). The *cat*RAPID predicted RNA-protein interactions of each viral RNA fragment were differently distributed within these intervals (Figures S6A and S6B). In particular, proteins above a $\log_2$ *cat*RAPID score of 5.22, corresponding with the 85th percentile and the graphical inflection point of the distribution, are significantly enriched for interactions with fragments 10, 4, and 5 (Figure 3C; Table S4). These three fragments together displayed 373 predicted interactions that were not found with the RNA control Scramble, including 13 interactions involving 9 RBPs significantly associated with the SARS-CoV-2 RNA fragments by RaPID-MS (Figure 3D). Among them were (i) FAM120A, LSG1, and RPS6 that were already reported as host proteins reproducibly interacting with SARS-CoV-2 RNA[24]; (ii) the splicing factors GPKOW, SF3A1, NOP58, and the NDM factor SMG7 that were computationally predicted to specifically associate to the genomic viral RNA[33,34]; and (iii) the RNA modifier PUS7, which catalyzes the isomerization of uridine into pseudouridine in cellular tRNAs and mRNAs, reported here for the first time. In the RaPID-MS dataset, PUS7 specifically associates with fragments 1, 4, 7, and 10, all harboring a PUS7 consensus sequence. Importantly, pseudouridine has been found as the most abundant modification in SARS-CoV-2 RNA[10] and another member of the pseudouridine synthase family, PUS1, has been reported to interact with SARS-CoV-2 RNA.[14] These observations strongly suggest that, in human cells, the formation of pseudouridine residues on the SARS-CoV-2 genome could be catalyzed by the activity of cellular RNA-independent pseudouridine synthases, such as PUS1 and PUS7.

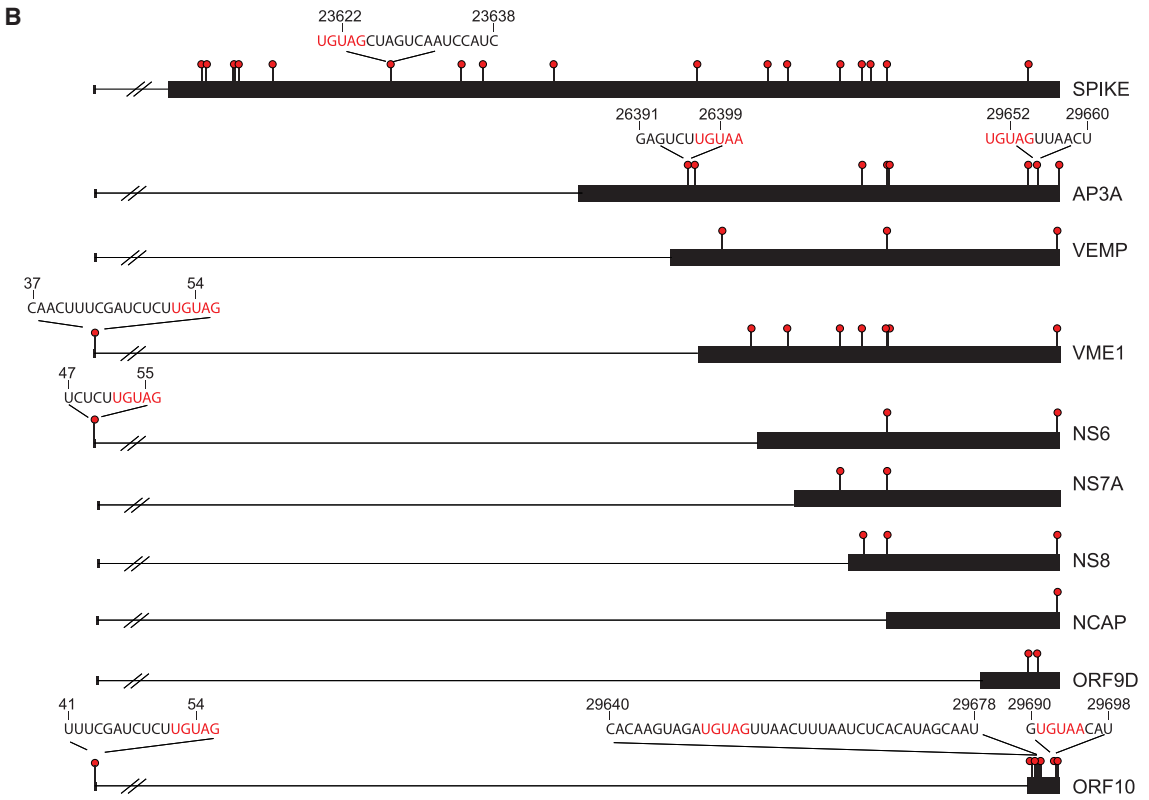
SARS-CoV-2 sgRNAs bear multiple putative pseudouridylated sites

To directly investigate occurrence of RNA modifications in SARS-CoV-2, including pseudouridine sites, we employed nanopore direct RNA sequencing (DRS) on cell lines infected with SARS-CoV-2 and analyzed the data with the Nanocompore software package (materials and methods). Nanocompore searches for RNA post-transcriptional modifications by comparing the ionic current features generated by nanopore DRS from two different experimental conditions: a test sample and a reference devoid of the modification of interest (or with a lower number of them).[35] Because of the physics of nanopore sequencing, Nanocompore compares the ionic current features of five nucleotides (named as k-mer) at a time. To identify putative pseudouridine sites, we filtered the results by selecting k-mers that contained at least one uridine and identified as significant by Nanocompore (p value $\leq$ 0.01; absolute value of the log odds ratio [LOR] $\geq$ 0.5). Initially, we used this technology to search for pseudouridine sites in the full-length genomic RNA (gRNA). We compared gRNA reads from CaCo-2, Calu-3, and Vero E6 infected with SARS-CoV-2 with a baseline reference, which consists of unmodified SARS-CoV-2 RNA obtained by *in vitro* transcription (IVT).[36] Nanocompore identified 63 significant k-mers (p value $\leq$ 0.01; absolute LOR $\geq$ 0.5), of which 58 having a uridine at the center of the identified sequence or in the first or second neighbor nucleotide (Figures 4, S7A, and S7B; Table S5). Among them, six k-mers were located within RaPID fragments 3–4, 5, 7–8, 8–9, and 9–10 (Table S5; Figure S7B).

Figure 3. *cat*RAPID performances on the RaPID-RBP datasets

(A) *cat*RAPID performances on the RaPID-RBP dataset. The predictive power of the method is calculated for different percentages of the dataset. *cat*RAPID performance is calculated on LFQ experimental value, normalized by taking into account the abundance of the proteins in PAXdb and protein length in Uniprot. (B) *cat*RAPID performance on the RaPID-RBP dataset, focusing on the single RNA fragments. For each fragment, the area under the curve at the different percentage of the dataset is shown. The performances are evaluated as in (A) and the fragments are ordered according to the respective genomic position. (C) (Left) Scatter chart of the analyzed RBPs ranked according to their relative $\log_2$ *cat*RAPID score. The dashed lines indicate the 85th percentile. (Right) Boxplot representation of the whole distribution divided in the indicated percentiles. (D) Scatter chart of the analyzed RBPs ranked according to their relative log2 *cat*RAPID score, displaying only proteins present in the 85th percentile of the distribution.





(legend on next page)

However, in the analyzed samples we noticed a coverage of only 140 reads spanning the whole 30-kb-long genome. Therefore, we extended our analysis to the sgRNAs that were more abundant, as previously quantified.[6] We analyzed all the 14 canonical sgRNAs of SARS-CoV-2 (Figure 4A; Tables S6, S7, and S8) used by the virus to translate the structural and accessory proteins required to produce new virions. In the attempt to characterize sgRNAs, we have previously employed Nanopore ReCappable sequencing (NRCeq), a new technique that can identify capped full-length RNAs.[6] Thus, we compared viral sgRNA from infected Vero E6, CaCo-2, and Calu-3 cells to the unmodified SARS-CoV-2 IVT RNA, used as a reference. Sequenced reads were aligned to the most up-to-date viral reference transcriptome[6] and the ionic current features from the RNA reads were realigned to each transcriptomic position of the reference using Nanopolish.[37] Then, we used Nanocompore to identify RNA post-transcriptional modifications as marked by differences in the electrical signal between the viral sgRNA reads and the IVT RNA reads (p value $\leq$ 0.01; absolute LOR $\geq$ 0.5). In total, we identified 1,164 (CaCo-2), 430 (Calu-3), and 627 (Vero E6) significant uridine-containing k-mers, distributed among the 14 canonical SARS-CoV-2 reference sgRNAs (Figures 4A and S8; Tables S6–S8).

To focus on sites consistently modified across samples, we considered only significant modified uridines that were identified in at least two out of three SARS-CoV-2-infected cell lines (see materials and methods). Overall, we obtained 457 candidate modified regions across the 10 different SARS-CoV-2 sgRNAs encoding for canonical ORFs (Figure 4B). Two of these regions were recently reported as a pseudouridylated sites in SARS-CoV-2 RNA[38] (Table S9). In addition, from a supplementary analysis comparing the significant uridine-modified k-mers found in the SARS-CoV-2 sgRNAs and gRNAs, we retrieved an overlap of 17 sites (Figure S7C; Table S5).

To shortlist the sites that have a greater likelihood of being pseudouridines (ψ) modified by PUS7, we selected those harboring the RNA consensus sequence of PUS7. We found that 53 sites had the more generic PUS7 consensus sequence UNUAR (red lollipops in Figure 4B), while eight had the more restrictive UGUAR motif (red lollipops with displayed sequence in Figure 4B). Six sites are located within RaPID fragments 1, 9, and 10, of which 1 and 10 were bound by PUS7 in our interactomic data (Table S9; Figure 2C). The modified sites containing the UGUAR consensus were manually inspected to confirm that the distributions of ionic current intensities and dwell times were different between the IVT and SARS-CoV-2 reads in a window of nine k-mers centered on the central U of the UGUAR motif (Figure S9). Interestingly, we observed that three of these sites were present within the TRS-L at the 5′ UTR of the sgRNAs encoding for ORF10, NS6, and VME1 (Figure 4B; Table S9).

PUS7-dependent pseudouridylation of TRS-L could modulate the binding of NSP1 to the viral 5′ UTR

In total, nanocompore analysis of the sgRNAs identified multiple putative post-transcriptionally modified sites in the TRS-L of SARS-CoV-2 (Figure 5A). Conversely, the UTR of the gRNA resulted devoid of any modification (Table S5). One of the modified sites is within the m6A consensus motif DRACH, which has recently been reported to be methylated by METTL3 and to regulate the translation rate of SARS-CoV-2 proteins.[39]

Another modified site is, instead, present in the PUS7 consensus sequence UGUAR, thus pointing toward a putative functional role of PUS7-dependent pseudouridylation of the SARS-CoV-2 UTR. We sought to understand the impact that the presence of a pseudouridine on the TRS-L might have. The modified site is present at the highly accessible uridine 54 (U54) within stem loop 2 (SL2) (Figure 5A), the most conserved region of the 5′ UTR of coronaviruses.[40] We obtained the chemically synthesized SL2 RNA sequence either unmodified or carrying a pseudouridylated U54. Structural investigation performed via circular dichroism revealed that the presence of ψ slightly increases the propensity of SL2 to form more stable secondary structures (Figure S10A), likely affecting the conformation of the SL2 pentaloop sequence.[41] Since RNA secondary structure is one of the drivers of the interactions with proteins, we reasoned that the presence of a pseudouridylated SL2 might favor its interaction with its main binder, the viral protein NSP1. NSP1 binds to the SARS-CoV-2 leader sequence to protect and promote the translation of viral RNA transcripts.[8,42,43] By means of biolayer interferometry, we estimated the impact of this pseudouridylation on the dissociation constant (K_d) with a recombinant NSP1. We observed that NSP1 can bind the *in vitro*-synthesized SL2 RNA sequence, with or without pseudouridylation, with a K_d in the high nanomolar range. However, the presence of ψ increased 2-fold the binding affinity, improving it from 300 to 170 nM (Figure S10B). The same trend of enhanced binding was assessed in cells through UV-RIP analysis performed with STREP-HA-tagged NSP1. We found that NSP1 is able to bind to the BoxB-RNA fragment 1 carrying the SARS-CoV-2 5′ UTR (Figure 5B). This RNA-protein interaction is enhanced by the ectopic expression of PUS7, while no differences were observed by the expression of a catalytic defective version of PUS7 (D294A)[44] (Figure 5B). These results hint at a possible role of PUS7-mediated pseudouridylation of SL2, a region included in the 5′ UTR that is typically bound by NSP1 to favor the translation of viral sgRNAs.

DISCUSSION

Host-virus interactions encrypt for the multiple processes of the viral life cycle, including, but not limited to, translation, transcription, and replication.[24,29,30] Unveiling the interactions between host proteins and both SARS-CoV-2 genomic and sgRNAs holds the potential

Figure 4. Nanocompore analysis identified significant k-mers in the analyzed cell lines

(A) UCSC Genome Browser (http://genome.ucsc.edu) annotation of the modified k-mers containing a uridine found in SARS-CoV-2 infected CaCo-2, Calu-3, and Vero E6 cells and distributed over 14 different reference sgRNAs. From the top to the bottom, IVT and RaPID fragments, SARS-CoV-2 NRCeq assembly and RefSeq SARS-CoV-2 ORFs are present as a reference. (B) Graphical representation of nucleotide ranges shared between at least two cell lines (red circles). Here, sites have been grouped per ORFs encoded by each reference sgRNA of the assembly and the sequence of those carrying an UGUAR motif is displayed.

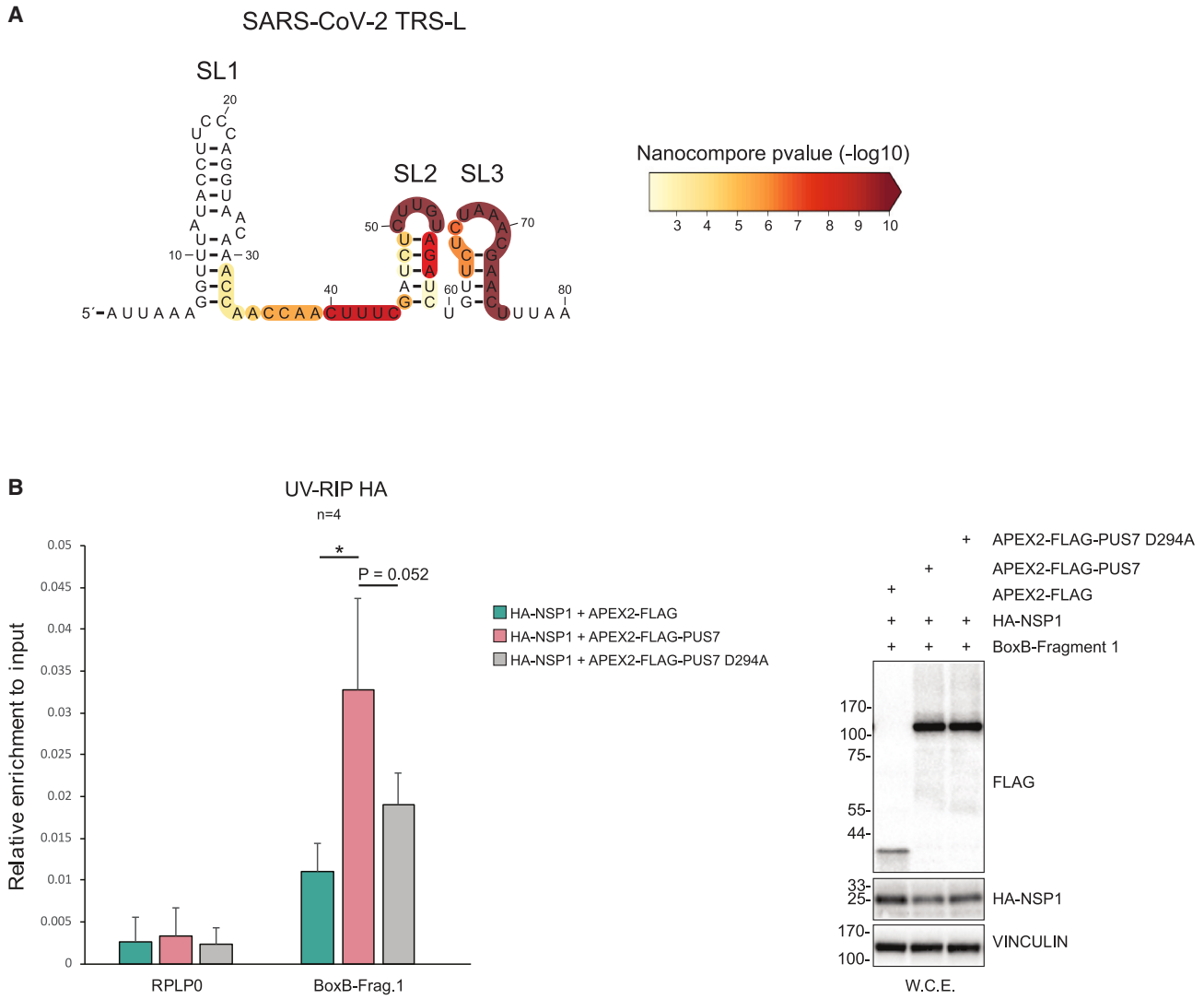


Figure 5. PUS7 modifies the SL2 of SARS-CoV-2 TRS-L favoring the binding of NSP1

(A) The secondary structure of SARS-CoV-2 TRS-L with the Nanocompore p value (NS6 sgRNA vs. IVT, GMM-logit test) overlaid as a color scale. The color represents the lowest p value among those of the five overlapping k-mers. Only k-mers with a p value ≤ 0.01 are colored according to the relative scale. (B) (Left) UV-RIP of HA-NSP1 co-expressed in HEK293 cells together with BoxB-Frag.1 expressing vector and either APEX2-FLAG, APEX2-FLAG-PUS7, or APEX2-FLAG-PUS7 D294A. The RNA bound to HA-NSP1 was eluted, retrotranscribed into cDNA and analyzed by RT-qPCR analysis using specific primers for RPLP0 and BoxB-Frag.1. The data shown are the average of four biological replicates (n = 4) and are expressed as relative enrichment over each respective input. Statistical analysis was performed using non-parametric, two-tailed t-test. *p value < 0.05. (Right) Representative image of the four WB analysis conducted on an aliquot of the UV-RIP input material (10%), showing the expression levels of the exogenous protein transiently co-expressed in HEK293 cells.

for the development of therapeutic strategies to inhibit the replication of this virus in human cells. Here, we propose an integrated approach that synergizes experimental results with computational predictions to construct a comprehensive portrait of the human proteins interacting with the SARS-CoV-2 RNA, enabling a better understanding of the molecular landscape of the virus-host interactome.

We exploited the state-of-the-art proximity ligation technology RaPID-MS,[20] a technique that enables the identification of the interactions between a known RNA sequence of interest and the proteins of the selected cell line. This method provides two critical advantages over other methodologies: (i) it circumvents any cross-linking step, reducing the number of proteins non-directly associated with SARS-CoV-2 RNA and (ii) it allows the identification of viral RNA-host protein interactions established during the early phases of infection. This critical period is marked by a minimal expression of viral proteins and displays the dependency of the virus on the interactions with the host proteins.

Existing studies have sought to determine the network of the SARS-CoV-2 RNA interactome with human proteins, yielding to a global view at various time points after infection.[12–15] However, a significant divergence is observed across these studies as a plausible consequence of different host cells and infection times selected, as well as of technical variability, such as the type of cross-linking agents employed to enrich for RNA-protein interactions.[24] In addition, these discrepancies underscore the complexity of the host-viral interface: inside infected cells, host proteins may engage simultaneously with the viral genomic RNA, subgenomic viral RNAs, or both, adding layers of complexity to these dynamics. By using our predictive model, *cat*RAPID, we determined that the host proteins most frequently reported across different studies are, indeed, predicted to have higher interaction propensity towards SARS-CoV-2 RNA, compared with the non-interacting RBPs.[29] This observation not only supports the strength of our predictions in identifying strong affinity interactions, but also forms the cornerstone of the strategy presented in this paper.

Our focus was primarily directed toward the most structured regions of SARS-CoV-2 RNAs, which are predicted to be highly contacted by human proteins.[18]

We identified several proteins (e.g., FAM120A, HAT1, LSG1, RPL14, RPL18A, RPL24, RPL35, RPS6 SHMT1, and SYNE2) that were previously uncovered by studies conducted on SARS-CoV-2-infected cells.[12–15] However, we noticed that some RBPs, reported to interact with SARS-CoV-2 RNA, were promiscuously interacting with our designed RNA fragments. Thus, these interactions likely contribute to response mechanisms triggered by the accumulation of viral RNA in the cell. The accumulation of viral RNA-host RBP interactions may necessitate the increase in local protein concentration offered by phase separation for the host proteins to sequester the viral RNA within liquid condensates.[18,29,45] In fact, different RBPs we identified as promiscuous binders, such as G3BP1, G3BP2, CAPRIN1, PUM1, and PUM2, are implicated in the formation and composition of stress granules.[45–47]

We also identified previously undisclosed SARS-CoV-2 RNA interactors, such as CEP350, GPKOW, and RGPD1, which, in our system, interact specifically with viral fragments containing the 5′ and 3′ UTR regions. These proteins seem to be key players in the genome cyclization process that potentially modulates SARS-CoV-2 RNA discontinuous transcription, as corroborated by recent reports.[5,32]

We were also able to link the interaction network of SARS-CoV-2 RNA with post-transcriptional modifications of the viral genomic and sgRNAs. Indeed, SARS-CoV-2 RNA is heavily post-transcriptionally modified,[48] although the role of each RNA modification has not been fully described. Modifications such as N6-methyladenosine, catalyzed by METTL3, have been associated with the modulation of the viral RNA sensor RIG-I, which effectively circumvents the host innate immune response.[10,49] Similarly, pseudouridine, which is documented to be the most abundant modification identified in SARS-CoV-2 RNA,[10] has been hypothesized to aid viruses to hijack the host immune sensors and evade the innate immune response.[50] The writer enzyme responsible for this modification is still unknown, although PUS1, a member of the pseudouridine synthase family, was found to be associated with SARS-CoV-2 RNA.[14]

Through our RaPID-MS analysis, we found PUS7, another member of the pseudouridine synthase family, binding to the RNA fragments present at both the 5′ and 3′ end regions of the viral genome. The retrieval of the PUS7 consensus sequence, UGUAR, within these fragments, paired with their differential electrical signal distribution compared to the unmodified *in vitro* transcribed RNA in nanopore DRS analysis, suggests PUS7 as the putative pseudouridylation agent for SARS-CoV-2 RNA.

Interestingly, we found three sgRNAs with a modified UGUAR site, as identified by Nanocompore, within the stem-loop 2 (SL2) of the SARS-CoV-2 TRS-L. This phenomenon is likely to occur in all SARS-CoV-2 sgRNAs, but it requires further improvements of the coverage at the 5′ end to be confidently scored by the nanopore DRS approach.

An important consequence of pseudouridylation is the potential effect on the RNA secondary structure and on its propensity to interact with both host and viral proteins. Specifically, we found evidence suggesting that PUS7-dependent pseudouridylation of the SARS-CoV-2 SL2 enhances the binding of the viral protein NSP1 to the viral 5′ UTR. An observation that fits the current model of NSP1-mediated regulation of the host translational machinery.[8,51,52] Indeed, the non-structural protein NSP1 is a well-known factor in manipulating the host translational machinery.[51,53] By binding to the 40S ribosomal subunit, NSP1 obstructs the mRNA entry site, effectively hindering translation of host mRNAs.[51,53] However, SARS-CoV-2 sgRNAs can elude NSP1-mediated translation inhibition thanks to the interaction of their 5′ UTRs with the N-terminal domain of NSP1, effectively dislodging the C-terminus from the ribosome entry channel.[43] It is plausible that PUS7 acts as a pro-viral factor, aiding the pseudouridylation of the 5′ UTR of sgRNAs to boost their translation rate. This hypothesis aligns with the recent proposal of NSP1 as a central element regulating RNA translation according to the affinity of each RNA toward NSP1.[52]

Both the sequence of SL2 and the NSP1-mediated regulation of translation are conserved in the sarbecovirus subgenus of betacoronaviruses, including SARS-CoV-1 and SARS-CoV-2.[54,55] This suggests that these viruses employ a shared strategy to prioritize the translation of their RNA transcripts over host mRNAs. Thus, a comprehensive understanding of the molecular mechanisms regulating the pseudouridylation of SARS-CoV-2 RNA could prove pivotal in devising innovative and targeted strategies to specifically impair the replication of sarbecoviruses in infected mammalian cells. In the future, the impairment of PUS7 activity could be exploited to interfere with SARS-CoV-2 replication. A PUS7 inhibitor, C17, has been recently described.[56] Alternatively, RNA molecules specifically targeting the

SL2 sequence can be designed to eventually inhibit its pseudouridylation with a more specific and effective approach.

Overall, our study has unveiled that even a single occurrence of pseudouridylation can potentially alter the critical interplay between the virus and host machinery, instrumental for the biology of the virus. It is conceivable that these effects are exponentially amplified when considering the potential cumulative impact of additional pseudouridylation sites spread across the viral RNA. Moreover, other types of post-transcriptional modifications to the viral RNA, such as methylation or acetylation, could yield similar or even more drastic effects on the virus-host dynamics. As such, the interplay between the virus and its host can be viewed as a complex and refined landscape, with each post-transcriptional modification representing a potential point of intervention.

This captivating field of study has the potential for ground-breaking discoveries that could profoundly reshape our understanding of viral biology. The advent of innovative technologies, such as nanopore DRS and related methods, is opening new frontiers for exploring this terrain. These technologies offer the potential to sequence RNA molecules in real time, providing unprecedented insights into the dynamics of RNA modifications. The capacity to probe the RNA molecules at the single-nucleotide level will enable us to map comprehensively these modifications across entire genomes. This could unlock new strategies to impede viral replication, either by disrupting the modification process itself or by altering its downstream effects. We foresee that this extensive understanding of RNA modification in SARS-CoV-2 and other viruses could modernize and improve our approach to antiviral therapies, transforming the ways to fight viral diseases.

MATERIALS AND METHODS

Preliminary predictions of protein-RNA interactions

The *cat*RAPID algorithm[25,57] was used to identify the binding propensity of the 10 SARS-CoV-2 500-bp fragments against a library of 2,064 human RBPs. For each fragment, only interactions with associated propensity score >85th percentile of the fragment score distribution were retained and subsequently, the average score was calculated.

Reagents and plasmids

The following plasmids were purchased from Addgene and used for the RaPID-MS assay: BoxB-plasmid (Addgene #107253); BoxB-EDEN15 plasmid (#107252); BASU RaPID plasmid (#107250). The following plasmids were a gift from Paul Khavari (Addgene plasmid #107253; #107252; #107250. http://n2t.net/addgene: 50917; 107252; 107250. RRID: Addgene_50917; Addgene 107252; Addgene 107250) (Ramanathan et al., 2018[20]). SARS-CoV-2 RNA fragments, derived from the SARS-CoV-2 RNA sequence MN908947.3, were synthesized by GeneArt (Thermo Fisher Scientific) according to the sequence reported in Table S1, flanked by Esp3I restriction sites. The same strategy was adopted for the Scramble control sequence: 3′-CGTCTCCGCTTTCGACGACAATTTATAAAGACAGCGGTC GAGGGAAGATTTACGAGTTGAATCGAGATGCGCTGATTCG ACGCAGTGTCGCGTTGTGGTGAGGTAAATTGATAGGTGTAT TTTGCGAGATACAGTGATGAACACTTCATTAACAACATGAT TTATACGACGATTACTAGAATTATGAAAAATGAGTCATCTA CAAGCGCGTTTTTACATTGCCGTGGTTAATCGTAAGGATAG CACAGTTAACAGCGGACCCCGGCGGACTCGGCCCTATCTGA ACGAATTGAGCTCCGTTCGAAATATCTAGTGAATGACCCTC CCCACGTGCCTTGATAAGCCGTGGTATTTCGTATCATACAA GTTCCAGAAGGATGGTTCAACATAGTAGGGTACCGACTGGA TAGAACAAACTACTCATGTTTTCGCCGGGGGACGAACGGTA AGCTCCGCTGGGTTGACTTCTTGACCAAAGTATTTGGGTAT CCAAACAGTGCCGTTAACAGCCAAGCTAGAGACG-5′.

The RNA fragments were then cloned into the pLEX BoxB-plasmid using the Esp3I sites, as described in literature.[20]

pDONR207 SARS-CoV-2 NSP1 was a gift from Fritz Roth (Addgene plasmid # 141255; http://n2t.net/addgene:141255; RRID: Addgene_141255). NSP1 coding sequence was cloned into the mammalian expression pCDNA5-FRT/TO-2xSTREP-3xHA vector, gently provided by G. Superti Furga, through the Gateway cloning system (Invitrogen).

The human coding sequence of PUS7 (variant 1, C-ter tag) cloned into the pDONOR221 vector was purchased from DNASU (HsCD00867933). The sequence was mutagenized to obtain the N-ter tag canonical variant 2 through the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent, 200521) using the following DNA primers: 5′- GGGAAAAAGGCTTTGGCAAATCCAAGAA AACATTCTTGGCC- 3′ and 5′- GGCCAAGAATGTTTTCTTGGA TTTGCCAAAGCCTTTTTCCC- 3′. The same strategy was used to generate the catalytic inactive mutant PUS7 D294A, using the following primers: 5′-ATTCTCCTACATGGGAACCAAAGCTAAA AGGGCTATAACAGTTC-3′ and 5′-GAACTGTTATAGCCCTTT TAGCTTTGGTTCCCATGTAGGAGAAT-3′.

The resultant cDNA was sequence verified and through the Gateway cloning system cloned into the pDEST-cDNA5-FRT/TO-3*Flag-APEX2 N-term, which was a gift from Benjamin Blencowe (Addgene plasmid # 182925; http://n2t.net/addgene:182925; RRID: Addgene_182925).

The following antibodies were used in this study: streptavidin-horseradish peroxidase (Cat. #3999) from Cell Signaling Technology. Anti-vinculin (Clone H Vin 1 0.2 ML; Cat. #V9131) from Merck. Anti-HA-11 epitope tag, clone 16B12 from BioLegend (Cat #901501), and Anti-Flag-M2 (Merck, F3165).

Cell lines

HEK293T cells were grown in DMEM with glucose and L-glutamine (Lonza, BE-12-604Q) supplemented with 10% fetal bovine serum (FBS), tetracycline free (Euroclone, ECS01822), and 100 U/mL penicillin and streptomycin. CaCo-2 cell lines were cultivated in Minimum Essential Medium Eagle (MEM) (Merck, M4655) complemented

with 20% FBS tetracycline free, 2 mM L-glutamine (Lonza, BE17605E), 1 mM sodium pyruvate (Lonza, BE13115E), 0.1 mM not essential amino acids (Lonza, BE13114E), and 100 U/mL penicillin and streptomycin (Euroclone, ECB3001D).

Virus isolation and cell infection

SARS-CoV-2 virus was isolated from a mildly symptomatic SARS-CoV-2 infected patient, as described by Ugolini et al.[6] CaCo-2 cells were infected at 80% confluency into a 25-cm^2 tissue culture flask with SARS-CoV-2 at an 0.1 multiplicity of infection. After 1 h adsorption at 37°C, cells were washed with PBS, and further cultured at 37°C for 48 h with 4% FBS. After a PBS wash, enzymatic dissociation was performed for 4–6 min at 37°C in 1 mL TrypLE (Invitrogen), then cell pellets were washed with ice-cold PBS and lysed with 1 mL of TRIzol (Invitrogen). The samples were stored at −80°C for subsequent RNA extraction.

RNA extraction and nanopore DRS

RNA was extracted from CaCo-2 cells infected with SARS-CoV-2 using the RNeasy Mini kit (Qiagen) including an in-column DnaseI treatment step, following the manufacturer protocol. The isolated RNA was then processed by direct RNA sequencing protocol as described by Ugolini et al.[6]

RaPID assay

The RaPID protocol was performed as reported[20] and slightly adapted as described below. HEK293T cells were transfected with plasmid vectors expressing λN-HA-BASU and one of the BoxB-RNA fragments using Lipofectamine 3000 transfection reagent (L3000001; Thermo Fisher Scientific), according to vendor's instructions. After 48 h from transfection, medium was changed and replaced with standard growth medium complemented with 200 μM biotin (Merck, B4639-1G) for 1 h. Cells were harvested, washed once with PBS 1× and lysed with RIPA buffer (10 mM Tris-HCl pH 8.0, 1%Triton X-100, 150 mM NaCl, 0.1% SDS, 0.1% NaDeoxycholate, 1 mM EDTA) supplemented with 1 mM 1,4-dithiothreitol (DTT), cOmplete Protease Inhibitor Cocktail (11697498001; Merck), and PhosSTOP (4906845001; Merck). Cell lysates were incubated for 15 min on ice and then centrifuged at 15,000×*g* for 15 min. The supernatants containing the protein extracts were transferred into fresh 1.5-mL tubes and protein concentration was measured by Bio-Rad Protein Assay Kit using BSA as protein standard (5000002; Bio-Rad). From each sample, 3 mg protein extract was taken and brought to the same volume (600 μL) with RIPA buffer. Five percent of input material was taken for further analysis and 150 μL pre-washed Streptavidin Mag Sepharose (GE28-9857-99; Merck) were added to the remaining material. Then, samples were rocked over night at 4°C. The following day, beads were separated from the unbound fractions and 5% of each fraction was collected in fresh tubes. Beads containing the biotinylated proteins were washed 3 times with 1 mL of Wash Buffer 1 (1% SDS supplemented with 1 mM DTT, protease, and phosphatase cocktail inhibitors); 1 time with Wash Buffer 2 (0.1% Na-DOC, 1% Triton X-100, 0.5 M NaCl, 50 mM HEPES pH 7.5, 1 μM EDTA supplemented with 1 mM DTT, protease, and phosphatase cocktail inhibitors) and 1 time with Wash Buffer 3 (0.5% Na-DOC, 150 mM NaCl, 0.5% NP-40, 10 mM Tris-HCl, 1 μM EDTA supplemented with 1 mM DTT, protease, and phosphatase cocktail inhibitors). All the washes were performed by rocking samples for 5 min at 4°C. Finally, proteins were eluted with Laemmli buffer containing 100 mM DTT, boiled for 5 min at 95°C and processed by *in-gel* digestion. The resultant peptide mixtures were analyzed by nano-LC-MS/MS analysis.

In-gel digestion

Eluted biotinylated proteins were processed as previously described.[58] Briefly, proteins were initially separated on a precast 4%–12% gradient gel (NP0322BOX, ThermoFisher Scientific). Each lane was divided in six slices that were cut from gels and destained in 50% v/v acetonitrile (ACN)/50 mM NH_4HCO_3. A reduction step was performed with 10 mM DTT, followed by alkylation with 55 mM iodoacetamide in the dark. After each step, samples were dehydrated with 100% ethanol and quickly dried in a centrifugal evaporator (SpeedVac). Subsequently, gel pieces were washed with 50 mM NH_4HCO_3 and overnight digested with 12.5 ng/mL trypsin (Promega, V5113) at 37°C. The following day, tryptic digested peptides were extracted with Extraction Buffer (3% TFA, 30% ACN) and 100% ACN. Prior to MS, peptides were desalted and concentrated in a single step through reversed phase chromatography on microcolumn C18 Stage Tips[59] and eluted in 0.1% formic acid (FA).

Nano-LC-MS/MS analysis

Peptide mixtures were analyzed by online nano-flow LC-MS/MS using an EASY-nLC 1000 (Thermo Fisher Scientific) connected to a Q-Exactive Plus instrument (Thermo Fisher Scientific) through a nano-electrospray ion source. The nano-LC system was operated in one column set-up with a 50-cm analytical column (75 μm inner diameter, 350 μm outer diameter) packed with C18 resin (EasySpray PEPMAP RSLC C18 2 μm 50 cm × 75 μm, Thermo Fisher Scientific) configuration. Solvent A was 0.1% FA in water and solvent B was 0.1% FA in 80% ACN. Samples were injected in an aqueous 0.1% TFA solution at a flow rate of 500 nL/min and separated with a gradient of 5%–40% solvent B over 50 min followed by a gradient of 40%–60% for 10 min and 60%–80% over 5 min at a flow rate of 250 nL/min in the EASY-nLC 1000 system. The Q-Exactive was operated in the data-dependent mode to automatically switch between full scan MS and MSMS acquisition. Survey full scan MS spectra (from m/z 300–1150) were analyzed in the Orbitrap detector with resolution R = 35,000 at m/z 400. The 10 most intense peptide ions with charge states R 2 were sequentially isolated to a target value of 3e6 and fragmented by higher energy collision dissociation with a normalized collision energy setting of 25%. The maximum allowed ion accumulation times were 20 ms for full scans and 50 ms for MSMS and the target value for MSMS was set to 1e6. The dynamic exclusion time was set to 20 s.

Data analysis of MS data

Proteins were identified and quantified using MaxQuant software v.1.6.0.16. using the Andromeda search engine.[60,61] In MaxQuant,

the false discovery rate of all peptide identifications was set to a maximum of 1%. The main search was performed with a mass tolerance of 6 ppm. Enzyme specificity was set to Trypsin/P. A maximum of three missed cleavages was permitted and the minimum peptide length was fixed at seven amino acids. Carbamidomethylation of cysteines was set as a fixed modification. The 2021_01 version of the human UniProt reference proteome (UP000005640) was used for peptide identification. Proteins were profiled by quantitative label-free analysis,[62] activating the label-free software MaxLFQ[60] and analyzed using Perseus software,[63] plotting the LFQ values in a volcano plot graph, where the proteins enriched with each SARS-CoV-2 RNA fragments were compared with the Scramble RNA control. The p value was calculated by Perseus using a two-tailed t-test. Missing values were replaced by the minimum detection value of the matrix. Proteins found as significantly enriched also with the EDEN15 RNA were removed from the final network. The interaction network generated with the statistically significant proteins was visualized using Cytoscape 3.8.1.[64]

Computational evaluation of experimental interactions

The *cat*RAPID algorithm was used to assess the predictive accuracy on the experimental RAPID dataset. Here, for each SARS-CoV-2 fragment-human protein pair, a fragmentation procedure was performed to identify the binding regions. The propensity score is calculated taking into consideration the maximum and the minimum score of all the sub-fragments computed by the algorithm[25] and computing their difference.[65] The median LFQ value for each protein-fragment pair was then normalized taking into account protein abundance in HEK293 cell lines from PAXdb[66] and protein length from UniProt database,[67] using the following equation:

$$Normalized\ LFQ\ value\ =\ LFQ - 0.6 * log_2(abundance) - 0.0001 * log_2(length).$$

This approach was introduced to reduce the experimental bias in LFQ scores with high protein abundance and length.

GO analysis

The GO analysis was performed on the 73 proteins specifically associated with the 10 SARS-CoV-2 RNA fragments and identified by RaPID-MS using the GOTERM biological process and molecular function present in the DAVID 6.8 Bioinformatics Resources (https://david-d.ncifcrf.gov/home.jsp).[68]

RNA secondary structure investigation

Circular dichroism analyses were performed on the RNA oligonucleotide "SL2" of sequence 5′-AACCAACUUUCGAUCUCUUGUAGAUCUGUUCU-3′ synthesized by Eurofins, and on the pseudouridylated SL2 sequence 5′-AACCAACUUUCGAUCUCUUG(psi)UAGAUCUGUUCU-3′, synthesized by Tebu-bio. The two versions of SL2 were prepared in a water stock at 100 μM and diluted to a concentration of 20 μM in a buffer containing 20 mM potassium phosphate at pH 7.2 and 150 mM kF. The RNA was then incubated for 1 h at 30°C to allow correct folding. Data acquisition of near-UV CD spectra was performed in 1-mm path-length quartz cuvettes on a JASCO-1100 spectropolarimeter supplied with a constant N_2 flush at 3.0 L/min. The experiment was performed in triplicate.

Nanopore DRS and nanocompore analysis

The nine DRS datasets were compared with the IVT dataset, using the NRCeq assembly and the methods in.[6] The dataset of Vero E6 cells, Calu-3 and CaCo-2 samples 1 and 2 are derived from[6] (European Nucleotide Archive [ENA] ID: PRJEB48830); CaCo-2 sample 3 and 4 are deposited with this manuscript (ENA ID: PRJEB53497). Reads were grouped by cell type and their fastq files concatenated. Reads were resquiggled using F5C (v0.6).[69] IVT reads were separately mapped to each reference sgRNA using minimap2 (v2.17-r974-dirty)[70] with the following parameters: -ax map-ont -p 0 -N 10.

Nanocompore (v1.0.4)[35] was used to detect RNA modifications by comparing the IVT reads to a set of reads from each cell line for every reference sgRNA (Tables S5; S6; S7) with the subsequent parameters:

–fasta <NRCeq_assembly_fasta_file> –overwrite –downsample_high_coverage 5000 –allow_warnings –pvalue_thr 0.01 –min_coverage 30 –logit –nthreads 3 –bed <NRCeq_assembly_bed_file>.

Merging analysis

Individual significant k-mers from the different cell lines may be offset by one or two positions, therefore the significant k-mers were transformed into larger nucleotide ranges of nine nucleotides, that include the five nucleotides of the significant k-mer, plus the two neighboring nucleotides present at both its extremities. Any of these nine nucleotides may be a post-transcriptional modified nucleotide detected by Nanocompore. In addition, we minimized ambiguous mapping by counting significant k-mers found on any isoform of VME1 and NCAP as a single isoform according to their genomic coordinates. Sites were compared with the "high-confidence sites" defined by Fleming et al.[38] using the single-nucleotide genomic position. Sites were manually inspected by plotting the distribution of signal intensity and dwell time for each position using functions implemented in Nanocompore.[35]

Comparison with RaPID fragments

Both significant k-mers and sites were compared with the RaPID fragments. Genomic coordinates were used for all comparisons with RaPID fragments (Table S1). In the case of k-mers, we defined a RaPID fragment match when the first nucleotide of the k-mer overlapped with any RaPID fragment, while for sites, we searched for every nucleotide of the site that could overlap any fragment. Genomic regions overlapping two fragments were identified with both fragment names. K-mers were assigned to each fragment based on the genomic position of their first nucleotide.

Genomic RNAs analysis

All infected datasets were mapped to the viral genome reference with minimap2 with the subsequent parameters.

-k 8 -w 1 -ax splice -g 30000 -G 30000 -A1 -B2 -O2,24 -E1,0 -C0 -z 400,200 –no-end-flt -F 40000 -N 32 –splice-flank = no –max-chain-skip = 40 -un -p 0.7.

All the positive strand primary alignments, with no deletions, ranging from at least 45 and 29,850 in genomic coordinates were extracted from all the datasets. These reads were compared with the IVT reads, pre-aligned to the viral reference genome, through Nanocompore with the subsequent parameters.

–downsample_high_coverage 2100 –allow_warnings –min_coverage 10 –logit.

The downsampling parameter was set to 2,100 as it was obtained from the multiplication of the number of gRNA reads (140) for the number of IVT fragments (15), as downsampling is performed over all the dataset. K-mers were evaluated nucleotide per nucleotide, to assess their overlap with IVT junctions, RaPID fragments, Fleming et al. sites, sgRNAs sites from the previous analysis and SNPs. The list of SNPs collects data from the literature (see Table S5). Only k-mers not overlapping any SNP or IVT junction were considered. K-mers were considered significant if having an absolute LOR $\geq$0.5 and a Gaussian Mixture Model (GMM) p value $\leq$0.01. We then investigated whether, despite the low coverage noticed for gRNAs, a comparison between the sgRNAs-specific sites and the gRNAs-specific ones could be performed. Therefore, to make the comparison between the gRNAs and the sgRNAs datasets fair, we pulled together all the data from each cell line analyzed and selected sgRNAs uridine-containing significant k-mers present in at least one canonical transcript model (see Table S10). To determine if the common sites were randomly selected or if such a number was to be statistically expected, we performed a hypergeometric test on common sites located after the TRS-B spike junction sequence, both for gRNAs and sgRNAs. The significance of the test (p = 1.44e-15; see github directory) suggests that, if we increased the coverage of the gRNAs dataset, we might retrieve the same sites found for the sgRNAs in the region starting from the TRS-B spike junction sequence to the end of the viral genome. The test could not be performed on the region starting at the genomic zero coordinate to the TRS-B spike junction sequence, as, of course, sgRNAs do not cover this portion of the genome.

UV-RIP assay

UV-RIP protocol was modified from.[71] Briefly, HEK293T were seeded in 10-cm dishes and transfected with the indicated mammalian expression plasmids were harvested, washed once with PBS and UV-cross-linked on ice with two cycles of irradiations at 100,000 μJ/cm^2. Cells were lysed with Lysis Buffer (0.5% NP-40, 0.5% NaDeoxycholate, 1× Roche protease inhibitors mixture (04693116001 Merck), 25 U/mL RNAse inhibitor (M03070L, NEB)) in PBS and rocked on a wheel for 30 min at 4°C. Afterward, lysates were treated with 30 U of Turbo DNAseI (AM2239, ThermoFisher Scientific) for 30 min at 37°C in a Thermomixer rotating at 1,100 rpm and subsequently centrifuged for 5 min at 1,350×*g* to remove cellular debris. The supernatants were used for the immuno-purification experiment, with an aliquot (10%) kept as input material. One-half of the supernatant volume (45%) was incubated with 4 μg HA antibody (901502, BioLegend) in a final volume of 500 μL with additional Lysis Buffer and rocked overnight at 4°C. The day after, 50 μL protein G dynabeads (10007D, ThermoFisher Scientific) were first washed three times in Lysis Buffer and then added to each sample. Samples were rocked for additional 3 h at 4°C. Afterward, dynabeads were washed four times with Washing Buffer I (PBS supplemented with 1% NP-40, 0.5% NaDeoxycholate, 300 mM NaCl, 1× Roche protease inhibitor mix and 25 U/mL RNAse inhibitor). The dynabeads were then resuspended in 100 μL RNAse-free water and treated again with the Turbo DNAseI for 30 min at 37°C, in a Thermomixer at 1,100 rpm. Input material was also treated with DNAseI for a second time. The dynabeads were then washed four times with Washing Buffer II (PBS supplemented with 1% NP-40, 0.5% NaDeoxycholate, 300 mM NaCl, 10 mM EDTA, 1× Roche protease inhibitor and 25 U/mL RNAse inhibitor). Finally, the RNA was eluted from the beads using the RNA-Lysis Buffer (Zymo Research) and extracted through the RNA-extraction kit (Zymo Research). The RNA was retro-transcribed into cDNA using the ImProm-IITM (A3800, Promega), according to the vendor's instruction. The cDNA was then diluted 1:5 with water and 5% of the diluted material was analyzed by RT-qPCR analysis using the Fast SYBR green master mix (4385614, ThermoFisher Scientific). The values obtained for each immunoprecipitated RNAs were normalized over the respective input material and plotted in a histogram, as relative fold enrichment.

DATA AND CODE AVAILABILITY

- The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[72] partner repository. ProteomeXchange: PXD034941.
- The RNA sequencing datasets generated in this study have been deposited at the European Nucleotide Archive (ENA) ENA browser: PRJEB53497.
- Custom tracks with data generated in this study are available at figshare: https://doi.org/10.6084/m9.figshare.24183363.v1, https://doi.org/10.6084/m9.figshare.24183324.v1, https://doi.org/10.6084/m9.figshare.24183501.v1. All the ONT analyses are available at the Github directory: https://github.com/nicassiolab/SARS-pseudoU.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2023.102052.

ACKNOWLEDGMENTS

The authors would like to thank the "RNA initiative" at the Istituto Italiano di Tecnologia (IIT). We thank Lorenzo di Tucci e Marco Rabozzi for developing an accelerated version of Nanopolish for preliminary analysis. The research leading to these results has been supported by the European Research Council (RIBOMYLOME_309545 and ASTRA_855923) and the H2020 projects (IASIS_727658 and INFORE_825080) to G.G.T.; the Italian Association for Cancer

Research (AIRC) - project IG 2019 (ID. 22851) to F.N. A grant from National Center for Gene Therapy and Drugs based on RNA Technology (CN00000041, EPNRRCN3) supported by European Union – NextGenerationEU PNRR MUR - M4C2 to F.N. and G.G.T. E.Z. received funding from the MINDED fellowship of the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 754490. C.U. is a PhD student within the European School of Molecular Medicine (SEMM) C.U. was supported by The Estée Lauder Companies Italia fellowship from the AIRC (ID.28410). L.M. was supported by an EMBL-IIT postdoctoral fellowships (ETPOD) fellowship.

AUTHOR CONTRIBUTIONS

R.G. and E.Z. designed and carried out the experiments, analyzed the data and wrote the manuscript; C.U. carried out the nanopore data analysis and wrote the manuscript; A.V. carried out the *cat*RAPID analysis; T.L. and L.M. contributed to the nanopore data analysis and the first draft of the manuscript; M.D.O. helped with data analysis and figure display; B.G. contributed to carry out the experiments; E.C., M.CA., N.C., M.Cl., and N.M. performed and were responsible for the generation of SARS-CoV-2 infected cells; T.B. helped with the analysis of the MS data; S.G. contributed with funding acquisition and supervision; G.G.T. and F.N. conceptualize the paper, contributed to the experimental design, wrote the manuscript, and provided funding. All authors actively contributed to this work, revised the manuscript, and agreed to its content.

DECLARATION OF INTERESTS

L.M. has received financial support from ONT for travel and accommodations to attend and present at ONT events. T.L. is a paid consultant to STORM therapeutics limited. This research was conducted in the absence of any commercial or financial relationships.

REFERENCES

1. Mistry, P., Barmania, F., Mellet, J., Peta, K., Strydom, A., Viljoen, I.M., James, W., Gordon, S., and Pepper, M.S. (2021). SARS-CoV-2 Variants, Vaccines, and Host Immunity. Front. Immunol. *12*, 809244. https://doi.org/10.3389/fimmu.2021.809244.

2. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. *5*, 536–544. https://doi.org/10.1038/s41564-020-0695-z.

3. Harrison, A.G., Lin, T., and Wang, P. (2020). Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. Trends Immunol. *41*, 1100–1115. https://doi.org/10.1016/j.it.2020.10.004.

4. Zhang, Y.-Z., and Holmes, E.C. (2020). A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. Cell *181*, 223–227. https://doi.org/10.1016/j.cell.2020.03.035.

5. Malone, B., Urakova, N., Snijder, E.J., and Campbell, E.A. (2022). Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. Nat. Rev. Mol. Cell Biol. *23*, 21–39. https://doi.org/10.1038/s41580-021-00432-z.

6. Ugolini, C., Mulroney, L., Leger, A., Castelli, M., Criscuolo, E., Williamson, M.K., Davidson, A.D., Almuqrin, A., Giambruno, R., Jain, M., et al. (2022). Nanopore ReCappable sequencing maps SARS-CoV-2 5′ capping sites and provides new insights into the structure of sgRNAs. Nucleic Acids Res. *50*, 3475–3489. https://doi.org/10.1093/nar/gkac144.

7. Sola, I., Almazán, F., Zúñiga, S., and Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. Annu. Rev. Virol. *2*, 265–288. https://doi.org/10.1146/annurev-virology-100114-055218.

8. Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., et al. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. Cell *183*, 1325–1339.e21. https://doi.org/10.1016/j.cell.2020.10.004.

9. Yin, X., Riva, L., Pu, Y., Martin-Sancho, L., Kanamune, J., Yamamoto, Y., Sakai, K., Gotoh, S., Miorin, L., De Jesus, P.D., et al. (2021). MDA5 Governs the Innate Immune Response to SARS-CoV-2 in Lung Epithelial Cells. Cell Rep. *34*, 108628. https://doi.org/10.1016/j.celrep.2020.108628.

10. Li, N., Hui, H., Bray, B., Gonzalez, G.M., Zeller, M., Anderson, K.G., Knight, R., Smith, D., Wang, Y., Carlin, A.F., and Rana, T.M. (2021). METTL3 regulates viral m6A RNA modification and host cell innate immune responses during SARS-CoV-2 infection. Cell Rep. *35*, 109091. https://doi.org/10.1016/j.celrep.2021.109091.

11. Kim, Y.-M., and Shin, E.-C. (2021). Type I and III interferon responses in SARS-CoV-2 infection. Exp. Mol. Med. *53*, 750–760. https://doi.org/10.1038/s12276-021-00592-0.

12. Kamel, W., Noerenberg, M., Cerikan, B., Chen, H., Järvelin, A.I., Kammoun, M., Lee, J.Y., Shuai, N., Garcia-Moreno, M., Andrejeva, A., et al. (2021). Global analysis of protein-RNA interactions in SARS-CoV-2-infected cells reveals key regulators of infection. Mol. Cell *81*, 2851–2867.e7. https://doi.org/10.1016/j.molcel.2021.05.023.

13. Lee, S., Lee, Y.S., Choi, Y., Son, A., Park, Y., Lee, K.-M., Kim, J., Kim, J.-S., and Kim, V.N. (2021). The SARS-CoV-2 RNA interactome. Mol. Cell *81*, 2838–2850.e6. https://doi.org/10.1016/j.molcel.2021.04.022.

14. Flynn, R.A., Belk, J.A., Qi, Y., Yasumoto, Y., Wei, J., Alfajaro, M.M., Shi, Q., Mumbach, M.R., Limaye, A., DeWeirdt, P.C., et al. (2021). Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions. Cell *184*, 2394–2411.e16. https://doi.org/10.1016/j.cell.2021.03.012.

15. Schmidt, N., Lareau, C.A., Keshishian, H., Ganskih, S., Schneider, C., Hennig, T., Melanson, R., Werner, S., Wei, Y., Zimmer, M., et al. (2021). The SARS-CoV-2 RNA–protein interactome in infected human cells. Nat. Microbiol. *6*, 339–353. https://doi.org/10.1038/s41564-020-00846-z.

16. Lei, X., Dong, X., Ma, R., Wang, W., Xiao, X., Tian, Z., Wang, C., Wang, Y., Li, L., Ren, L., et al. (2020). Activation and evasion of type I interferon responses by SARS-CoV-2. Nat. Commun. *11*, 3810. https://doi.org/10.1038/s41467-020-17665-9.

17. Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J.Y.-C., Wang, H., Menachery, V.D., Rajsbaum, R., and Shi, P.-Y. (2020). Evasion of Type I Interferon by SARS-CoV-2. Cell Rep. *33*, 108234. https://doi.org/10.1016/j.celrep.2020.108234.

18. Vandelli, A., Monti, M., Milanetti, E., Armaos, A., Rupert, J., Zacco, E., Bechara, E., Delli Ponti, R., and Tartaglia, G.G. (2020). Structural analysis of SARS-CoV-2 genome and predictions of the human interactome. Nucleic Acids Res. *48*, 11270–11283. https://doi.org/10.1093/nar/gkaa864.

19. Giambruno, R., and Nicassio, F. (2022). Proximity-dependent biotinylation technologies for mapping RNA-protein interactions in live cells. Front. Mol. Biosci. *9*, 1062448. https://doi.org/10.3389/fmolb.2022.1062448.

20. Ramanathan, M., Majzoub, K., Rao, D.S., Neela, P.H., Zarnegar, B.J., Mondal, S., Roth, J.G., Gai, H., Kovalski, J.R., Siprashvili, Z., et al. (2018). RNA–protein interaction detection in living cells. Nat. Methods *15*, 207–212. https://doi.org/10.1038/nmeth.4601.

21. Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., Vabulas, R.M., and Tartaglia, G.G. (2019). RNA structure drives interaction with proteins. Nat. Commun. *10*, 3246. https://doi.org/10.1038/s41467-019-10923-5.

22. Cerase, A., Calabrese, J.M., and Tartaglia, G.G. (2022). Phase separation drives X-chromosome inactivation. Nat. Struct. Mol. Biol. *29*, 183–185. https://doi.org/10.1038/s41594-021-00697-0.

23. Brant, A.C., Tian, W., Majerciak, V., Yang, W., and Zheng, Z.-M. (2021). SARS-CoV-2: from its discovery to genome structure, transcription, and replication. Cell Biosci. *11*, 136. https://doi.org/10.1186/s13578-021-00643-z.

24. Baggen, J., Vanstreels, E., Jansen, S., and Daelemans, D. (2021). Cellular host factors for SARS-CoV-2 infection. Nat. Microbiol. *6*, 1219–1232. https://doi.org/10.1038/s41564-021-00958-0.

25. Armaos, A., Colantoni, A., Proietti, G., Rupert, J., and Tartaglia, G.G. (2021). *cat* RAPID *omics v2.0* : going deeper and wider in the prediction of protein–RNA interactions. Nucleic Acids Res. *49*, W72–W79. https://doi.org/10.1093/nar/gkab393.

26. Cid-Samper, F., Gelabert-Baldrich, M., Lang, B., Lorenzo-Gotor, N., Ponti, R.D., Severijnen, L.-A.W.F.M., Bolognesi, B., Gelpi, E., Hukema, R.K., Botta-Orfila, T., and Tartaglia, G.G. (2018). An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. Cell Rep. *25*, 3422–3434.e7. https://doi.org/10.1016/j.celrep.2018.11.076.

27. Agostini, F., Cirillo, D., Bolognesi, B., and Tartaglia, G.G. (2013). X-inactivation: quantitative predictions of protein interactions in the Xist network. Nucleic Acids Res. *41*, e31. https://doi.org/10.1093/nar/gks968.

28. Armaos, A., Zacco, E., Sanchez de Groot, N., and Tartaglia, G.G. (2021). RNA-protein interactions: Central players in coordination of regulatory networks. Bioessays *43*, 2000118. https://doi.org/10.1002/bies.202000118.

29. Vandelli, A., Vocino, G., and Tartaglia, G.G. (2022). Phase Separation Drives SARS-CoV-2 Replication: A Hypothesis. Front. Mol. Biosci. *9*, 893067. https://doi.org/10.3389/fmolb.2022.893067.

30. Iselin, L., Palmalux, N., Kamel, W., Simmonds, P., Mohammed, S., and Castello, A. (2022). Uncovering viral RNA–host cell interactions on a proteome-wide scale. Trends Biochem. Sci. *47*, 23–38. https://doi.org/10.1016/j.tibs.2021.08.002.

31. Edwards, J.M., Long, J., de Moor, C.H., Emsley, J., and Searle, M.S. (2013). Structural insights into the targeting of mRNA GU-rich elements by the three RRMs of CELF1. Nucleic Acids Res. *41*, 7153–7166. https://doi.org/10.1093/nar/gkt470.

32. Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F., and Miska, E.A. (2020). The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. Mol. Cell *80*, 1067–1077.e5. https://doi.org/10.1016/j.molcel.2020.11.004.

33. Horlacher, M., Oleshko, S., Hu, Y., Ghanbari, M., Cantini, G., Schinke, P., Vergara, E.E., Bittner, F., Mueller, N.S., Ohler, U., et al. (2021). Computational Mapping of the Human-SARS-CoV-2 Protein-RNA Interactome. Bioinformatics. https://doi.org/10.1101/2021.12.22.472458.

34. Akbari, B., Ahmadi, E., Zamir, M.R., Shaker, M.S., and Noorbakhsh, F. (2022). SARS-CoV-2 Helicase might interfere with cellular nonsense-mediated RNA decay, insights from a bioinformatics study. Bioinformatics. https://doi.org/10.1101/2022.05.30.494036.

35. Leger, A., Amaral, P.P., Pandolfini, L., Capitanchik, C., Capraro, F., Miano, V., Migliori, V., Toolan-Kerr, P., Sideri, T., Enright, A.J., et al. (2021). RNA modifications detection by comparative Nanopore direct RNA sequencing. Nat. Commun. *12*, 7198. https://doi.org/10.1038/s41467-021-27393-3.

36. Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. Cell *181*, 914–921.e10. https://doi.org/10.1016/j.cell.2020.04.011.

37. Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods *14*, 407–410. https://doi.org/10.1038/nmeth.4184.

38. Fleming, A.M., Mathewson, N.J., Howpay Manage, S.A., and Burrows, C.J. (2021). Nanopore Dwell Time Analysis Permits Sequencing and Conformational Assignment of Pseudouridine in SARS-CoV-2. ACS Cent. Sci. *7*, 1707–1717. https://doi.org/10.1021/acscentsci.1c00788.

39. Aly, A., Scott, G., Calderon, M., and Haghighi, A.P. (2022). N6-Adenosine Methylation of SARS-CoV-2 5'-UTR Regulates Translation. Mol. Biol. https://doi.org/10.1101/2022.10.17.512569.

40. Miao, Z., Tidu, A., Eriani, G., and Martin, F. (2021). Secondary structure of the SARS-CoV-2 5'-UTR. RNA Biol. *18*, 447–456. https://doi.org/10.1080/15476286.2020.1814556.

41. Liu, P., Li, L., Keane, S.C., Yang, D., Leibowitz, J.L., and Giedroc, D.P. (2009). Mouse Hepatitis Virus Stem-Loop 2 Adopts a uYNMG(U)a-Like Tetraloop Structure That Is Highly Functionally Tolerant of Base Substitutions. J. Virol. *83*, 12084–12093. https://doi.org/10.1128/JVI.00915-09.

42. Vankadari, N., Jeyasankar, N.N., and Lopes, W.J. (2020). Structure of the SARS-CoV-2 Nsp1/5′-Untranslated Region Complex and Implications for Potential Therapeutic Targets, a Vaccine, and Virulence. J. Phys. Chem. Lett. *11*, 9659–9668. https://doi.org/10.1021/acs.jpclett.0c02818.

43. Mendez, A.S., Ly, M., González-Sánchez, A.M., Hartenian, E., Ingolia, N.T., Cate, J.H., and Glaunsinger, B.A. (2021). The N-terminal domain of SARS-CoV-2 nsp1 plays key roles in suppression of cellular gene expression and preservation of viral gene expression. Cell Rep. *37*, 109841. https://doi.org/10.1016/j.celrep.2021.109841.

44. Guzzi, N., Cieśla, M., Ngoc, P.C.T., Lang, S., Arora, S., Dimitriou, M., Pimková, K., Sommarin, M.N.E., Munita, R., Lubas, M., et al. (2018). Pseudouridylation of tRNA-Derived Fragments Steers Translational Control in Stem Cells. Cell *173*, 1204–1216.e26. https://doi.org/10.1016/j.cell.2018.03.008.

45. Onomoto, K., Yoneyama, M., Fung, G., Kato, H., and Fujita, T. (2014). Antiviral innate immunity and stress granule responses. Trends Immunol. *35*, 420–428. https://doi.org/10.1016/j.it.2014.07.006.

46. Jayabalan, A.K., Griffin, D.E., and Leung, A.K.L. (2023). Pro-Viral and Anti-Viral Roles of the RNA-Binding Protein G3BP1. Viruses *15*, 449. https://doi.org/10.3390/v15020449.

47. Narita, R., Takahasi, K., Murakami, E., Hirano, E., Yamamoto, S.P., Yoneyama, M., Kato, H., and Fujita, T. (2014). A Novel Function of Human Pumilio Proteins in Cytoplasmic Sensing of Viral Infection. PLoS Pathog. *10*, e1004417. https://doi.org/10.1371/journal.ppat.1004417.

48. Izadpanah, A., Rappaport, J., and Datta, P.K. (2022). Epitranscriptomics of SARS-CoV-2 Infection. Front. Cell Dev. Biol. *10*, 849298. https://doi.org/10.3389/fcell.2022.849298.

49. Burgess, H.M., Depledge, D.P., Thompson, L., Srinivas, K.P., Grande, R.C., Vink, E.I., Abebe, J.S., Blackaby, W.P., Hendrick, A., Albertella, M.R., et al. (2021). Targeting the m ^{6} A RNA modification pathway blocks SARS-CoV-2 and HCoV-OC43 replication. Genes Dev. *35*, 1005–1019. https://doi.org/10.1101/gad.348320.121.

50. Li, N., and Rana, T.M. (2022). Regulation of antiviral innate immunity by chemical modification of viral RNA. WIREs RNA *13*, e1720. https://doi.org/10.1002/wrna.1720.

51. Simeoni, M., Cavinato, T., Rodriguez, D., and Gatfield, D. (2021). I(nsp1)ecting SARS-CoV-2–ribosome interactions. Commun. Biol. *4*, 715. https://doi.org/10.1038/s42003-021-02265-0.

52. Tidu, A., Janvier, A., Schaeffer, L., Sosnowski, P., Kuhn, L., Hammann, P., Westhof, E., Eriani, G., and Martin, F. (2020). The viral protein NSP1 acts as a ribosome gatekeeper for shutting down host translation and fostering SARS-CoV-2 translation. RNA *27*, 253–264. https://doi.org/10.1261/rna.078121.120.

53. Schubert, K., Karousis, E.D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.-A., Leibundgut, M., Thiel, V., Mühlemann, O., and Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. Nat. Struct. Mol. Biol. *27*, 959–966. https://doi.org/10.1038/s41594-020-0511-8.

54. Sosnowski, P., Tidu, A., Eriani, G., Westhof, E., and Martin, F. (2022). Correlated sequence signatures are present within the genomic 5′UTR RNA and NSP1 protein in coronaviruses. RNA *28*, 729–741. https://doi.org/10.1261/rna.078972.121.

55. Yuan, S., Balaji, S., Lomakin, I.B., and Xiong, Y. (2021). Coronavirus Nsp1: Immune Response Suppression and Protein Expression Inhibition. Front. Microbiol. *12*, 752214. https://doi.org/10.3389/fmicb.2021.752214.

56. Cui, Q., Yin, K., Zhang, X., Ye, P., Chen, X., Chao, J., Meng, H., Wei, J., Roeth, D., Li, L., et al. (2021). Targeting PUS7 suppresses tRNA pseudouridylation and glioblastoma tumorigenesis. Nat. Cancer *2*, 932–949. https://doi.org/10.1038/s43018-021-00238-0.

57. Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G.G. (2011). Predicting protein associations with long noncoding RNAs. Nat. Methods *8*, 444–445. https://doi.org/10.1038/nmeth.1611.

58. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. *1*, 2856–2860. https://doi.org/10.1038/nprot.2006.468.

59. Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906. https://doi.org/10.1038/nprot.2007.261.

60. Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526. https://doi.org/10.1074/mcp.M113.031591.

61. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372. https://doi.org/10.1038/nbt.1511.

62. Giambruno, R., Mihailovich, M., and Bonaldi, T. (2018). Mass Spectrometry-Based Proteomics to Unveil the Non-coding RNA World. Front. Mol. Biosci. *5*, 90. https://doi.org/10.3389/fmolb.2018.00090.

63. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740. https://doi.org/10.1038/nmeth.3901.

64. Su, G., Morris, J.H., Demchak, B., and Bader, G.D. (2014). Biological Network Exploration with Cytoscape 3. Curr. Protoc. Bioinformatics *47*, 8.13.1–8.13.24. https://doi.org/10.1002/0471250953.bi0813s47.

65. Lang, B., Armaos, A., and Tartaglia, G.G. (2019). RNAct: Protein–RNA interaction predictions for model organisms with supporting experimental data. Nucleic Acids Res. *47*, D601–D606. https://doi.org/10.1093/nar/gky967.

66. Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics *15*, 3163–3168. https://doi.org/10.1002/pmic.201400441.

67. UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., et al. (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. *49*, D480–D489. https://doi.org/10.1093/nar/gkaa1100.

68. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57. https://doi.org/10.1038/nprot.2008.211.

69. Gamaarachchi, H., Lam, C.W., Jayatilaka, G., Samarakoon, H., Simpson, J.T., Smith, M.A., and Parameswaran, S. (2020). GPU accelerated adaptive banded event alignment for rapid comparative nanopore signal analysis. BMC Bioinf. *21*, 343. https://doi.org/10.1186/s12859-020-03697-x.

70. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

71. Jeon, Y., and Lee, J.T. (2011). YY1 Tethers Xist RNA to the Inactive X Nucleation Center. Cell *146*, 119–133. https://doi.org/10.1016/j.cell.2011.06.026.

72. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552. https://doi.org/10.1093/nar/gkab1038.