Multi-Modal SARS-CoV-2 Proteins Interactomics Unveil Spatial Patterns of Invasion Featuring 151 Contacts with Human Proteins Essential for Infection

Guillaume Dugied, Estelle Mn Laurent, Mikaël Attia, Jean-Pascal Gimeno, Kamel Bachiri, Payman Samavarchi-Tehrani, Flora Donati, Yannis Rahou, Sandie Munier, Faustine Amara, et al.

Multi-modal SARS-CoV-2 proteins interactomics unveil spatial patterns of invasion featuring 151 contacts with human proteins essential for infection

Guillaume Dugied[1,#,], Estelle MN Laurent[2,*], Mikaël Attia[1,#,*], Jean-Pascal Gimeno[2], Kamel Bachiri[2], Payman Samavarchi-Tehrani[3], Flora Donati[4], Yannis Rahou[1,4], Sandie Munier, Faustine Amara[1], Mélanie Dos Santos[1], Nicolas Soler[5], Stevenn Volant[6], Nathalia Pietrosemoli[6], Anne-Claude Gingras[3], Georgios A. Pavlopoulos[7], Sylvie van der Werf[1,4], Pascal Falter-Braun[8,9], Patrick Aloy[5,10], Yves Jacob[1], Anastassia Komarova[1#], Yorgos Sofianatos[7§✉], Caroline Demeret[1#§✉], Etienne Coyaud[2§✉].*

1 Institut Pasteur, Université Paris Cité, UMR 3569, Centre National de la Recherche Scientifique, Molecular Genetics of RNA Viruses, 28 rue du Docteur Roux, F-75015 Paris, France

2 Univ. Lille, Inserm, CHU Lille, U1192 - Protéomique Réponse Inflammatoire Spectrométrie de Masse - PRISM, F-59000 Lille, France

3 Lunenfeld-Tanenbaum Research Institute at Mount Sinai Hospital, Sinai Health, Toronto, Ontario, Canada

4 Institut Pasteur, Université Paris Cité, National Reference Center for respiratory viruses, 28 rue du Docteur Roux, F-75015 Paris, France

5 Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10 -12,08020 Barcelona, Spain.

6 Bioinformatics and Biostatistics Hub, Institut Pasteur, Université Paris Cité, F-75015 Paris, France

7 Institute for Fundamental Biomedical Research, BSRC "Alexander Fleming", 34 Fleming Street, 16672, Vari, Greece

8 Institute of Network Biology (INET), Molecular Targets and Therapeutics Center (MTTC), Helmholtz Center Munich, German Research Center for Environmental Health, Munich-Neuherberg, Germany

9 Microbe-Host Interactions, Faculty of Biology, Ludwig-Maximilians-Universität (LMU) München, Planegg-Martinsried, Germany.

10 Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys, 23, 08010 Barcelona, Spain.

** contributed equally to this work*

§co-last authors

✉ co-corresponding authors

#present address: : Institut Pasteur, Université Paris Cité, Interactomics, RNA and Immunity, 28 rue du Docteur Roux, F-75015 Paris, France.

Summary

From a comprehensive proximity-dependent biotinylation (BioID)-based proximal SARS-CoV-2 interactome, we generated a 3D map positioning both SARS-CoV-2 and host proteins within cell volumes. Then, a subset of the host proteome proximal to the viral polypeptides was probed for direct contacts by a semi-quantitative split-nanoluciferase assay. We thus provide an intensity map of 414 direct virus-host protein contacts, of which 151 involved host proteins essential for SARS-CoV-2 infection. Among the 70 virus-host complexes modeled by AlphaFold2, 10 scored positive, providing as many candidate interfaces relevant for structure-driven screening of PPIs-disrupting compounds. Based on our data, we selected and showed that USP13 inhibition potently inhibits SARS-CoV-2 replication. This innovative and multi-pronged pipeline provides a resource of unprecedented depth. Its capacity in identifying virus-host PPIs highly relevant for the screening of antiviral compounds and its versatility for any emerging human pathogens could make this pipeline a prime tool for viral preparedness.

Keywords: Protein-Protein Interactions, SARS-CoV-2, Systems Biology, BioID, split-nanoluciferase, Alphafold2, 3D interactive topology.

Introduction

The COVID-19 pandemic revealed the need to improve our reactivity in developing effective countermeasures when facing emerging pathogens. Controlling the spread of a pathogen in an immunologically naive population requires prevention, containment, as well as the timely implementation of therapeutics to limit transmission and severe health effects. In addition, after over three years, the COVID-19 pandemic persists as a serious public health concern, especially in countries with low vaccination rates. Considering the series of COVID-19 waves due to the emergence of SARS-CoV-2 variants, the control of COVID-19 disease remains challenging. Further characterization of SARS-CoV-2 pathogenesis at a molecular level is still needed to combat viral mechanisms through antiviral therapeutics. Identifying host proteins to target is necessary for long-term COVID-19 management due to their lower mutation rate compared to the viral ones. Therefore, determining virus-host protein interfaces and developing targeted therapies to disrupt their contact appears promising for host-targeted therapeutics.

Here, we describe an interactomics pipeline that we applied to SARS-CoV-2 and that: (**i**) can be deployed as soon as the genome of any pathogen is sequenced; (**ii**) identifies host-pathogen promiscuity within cellular

territories; (**iii**) identifies direct pathogen/host contacts possibly essential for infection; and (**iv**), provide protein-protein interaction (PPI) interfaces of strong therapeutic interest.

SARS-CoV-2 is a positive-sense RNA virus replicating in the cytoplasm and strongly depending on intracellular membranous organelles for its productive life cycle. Nine of the twenty-nine viral proteins are localized in intracellular organelles and exhibit transmembrane domains[1,2]. Thus, membranes- and organelles-adapted biochemical protocols are best suited to study SARS-CoV-2 proteins interactions within the host cell.

Proteomic studies relying on classical affinity purification methods followed by mass spectrometry (AP-MS)-based identification of protein complexes have been performed to define the interplay of SARS-CoV-2 proteins with host cell components[3–6]. One major limitation of AP-MS is the requirement of maintaining protein-protein interactions (PPIs) throughout the purification process, which necessitates the use of a gentle lysis buffer to solubilize physically associated proteins without disrupting PPIs, hence precluding the detection of PPIs within poorly soluble compartments. The proximity-dependent biotinylation technique (BioID)[7] allows the identification of proximal PPIs of a given $BirA_{R118G}$(BirA*)-tagged protein in living cells through their covalent biotin labeling, eliminating the requirement for the preservation of PPIs upon verification process. Therefore, proximal partners of the BirA*-tagged bait proteins are captured regardless of their solubility and PPI affinity. Since SARS-CoV-2 heavily relies on membranous organelle-associated processes, proximal interactomics pipelines, such as BioID, are best suited to explore the protein environment in which SARS-CoV-2 polypeptides operate to hijack host cells[2]. We used BioID-based proximal interactomic approaches to delineate the interplay of SARS-CoV-2/human proteins (**Fig.S1**) and better understand the invasion features of SARS-CoV-2 within cells. This dataset allowed us to construct a 3D map of the approximate relative localizations of virus and host proteins within the cell volume (**Fig.S1**). Featuring a user-friendly interface, this 3D map is an invaluable tool for unraveling the SARS-CoV-2-mediated hijacking of specific regions of the host proteome.

Orthogonal approaches detecting binary PPIs provide insight on the physical contacts between viral and host proteins, which are underlying virus-driven hijacking of the host proteome. Therefore, we assessed direct virus-host protein contacts by implementing a split-NanoLuciferase protein complementation assay performed in mammalian cells (mN2H)[8]. PPI scoring based on interaction-mediated luciferase signals gave rise to interaction maps highlighting the prevailing virus-host contacts among proximal host proteins (**Fig.S1**). The systematic exploration of direct virus-host contacts among virus-proximal host proteins revealed 414 direct interactions between 26 viral and 118 host proteins. Intriguingly, 41 of these host proteins (35%) have been reported as critical for SARS-CoV-2 infection in 3+ CRISPR-Cas9 screens, corroborating the central role of direct virus interactors for SARS-CoV-2 infection and replication[9,10]. Subjecting several virus-host complexes to *in silico*

structure prediction using AlphaFold2 Multimer[11] or structure-based homology (**Fig.S1**), we obtained multiple high confidence structures. Moreover, using an in-house assay for sensitive and quantitative detection of SARS-CoV-2 infection in cell lines, we confirmed the potential of USP13 chemical inhibition to combat SARS-CoV-2 replication (**Fig.S1**), thus providing a strong rationale for a structure-directed search of USP13-NSP13 inhibitors.

Overall, our dataset templates the first interactive 3D modeling of the viral-host proteomes interface in the cell volume, providing an unparalleled picture of the SARS-CoV-2 invasion pattern. Our pipeline culminates in identifying high-confidence virus-host protein interfaces, paving the way for identifying PPI-disrupting compounds with high therapeutic potential. Significantly, being applicable to any virus or intracellular pathogen as soon as the genome is sequenced, this pipeline could represent a major asset for responding rapidly to emerging pathogens and improving preparedness efficiency.

Results

The landscape of SARS-CoV-2 proteins' cellular environment and physical interactions cores

We used BioID to capture the neighbors of 27 SARS-CoV-2 polypeptides: (**i**) 15 of the 16 non-structural proteins that are produced in infected cells from the proteolytic cleavage of the Orf1a/b polyproteins. NSP11 was excluded due to its small size (12 AAs); (**ii**) the four structural proteins: Spike (S) glycoprotein, Envelope (E) protein, Membrane (M) glycoprotein, and Nucleocapsid (N) protein; (**iii**) eight accessory proteins: Orf3a, Orf3d, Orf6, Orf7a, Orf7b, Orf8, Orf9b, Orf9c (**Fig.S1** and **supp table 1**). Following biochemical assays showing that the BirA*-viral bait proteins phenocopy their untagged counterparts in impairing the innate immune responses (**Fig.S2-S4**), we proceeded to the identification of proximal host factors of the 27 SARS-CoV-2 proteins. Our experiments were performed either in basal conditions (n=3 with N- and C-ter BirA* tag) or following poly(I:C) induction (n=3) to mimic a cellular antiviral response and subsequent host cell proteome remodeling. A total of 243 SARS-CoV-2 exploratory BioID samples were generated, and the BioID background was determined with 30 negative control runs performed using cells expressing BirA* alone, covering all the experimental settings. High confidence interactors (*p*-values <0.01) were defined based on fold change (FC) over background and detection in three experimental replicates. Aggregating all identified neighbors across the different replicates and conditions, we obtained 10,125 high confidence proximal interactions between the 27 SARS-CoV-2 proteins and 2,593 host partners.

To benchmark our proximal interactomics dataset, we compared our results to all previously reported SARS-

CoV-2 interactions data in the BioGRID database (https://thebiogrid.org/). We observed a high overlap with the proximal interactomic datasets even though the experimental conditions vary substantially with regard to the used systems (BioID, TurboID and BioID2), cell lines (HEK293 and A549), purification protocols and analytical tools. In total, 4,738/10,125 (47%) of our PPIs have been identified in at least one other proximal interactomic study. Conversely, the percentages of the reported proximal interactions that are present in our dataset are 81.4% (2903/3565; BioID in HEK293 address only part of the SARS CoVo2 proteome[12]), 38.5% (3003/7810; TurboID in A549[13]) and 24.7% (172/695 BioID2 in A549[14]). In addition, 1089 of our 10,125 proximal PPIs were recovered in the body of SARS-CoV-2 viral proteins interactions studies obtained by different methods, *i.e.* AP-MS, AP-WB, Y2H, crystallography (**supp table 2**). Combining our SARS-CoV-2 interactome datasets with the existing ones, two validated core datasets can be distinguished (**supp table 3**): (**i**) a proximal core, supported by other proximity-dependent biotin labeling studies, encompassing 4696 proximal interactions between 27 viral bait proteins and 1,361 host proteins; and (**ii**), a physical core, *i.e.* the interactions confirmed by 1+ other method detecting proteins complexes, including direct and indirect PPIs. The physical core encompasses 1088 interactions between 659 host proteins and 27 viral bait proteins (**Fig.1A**). Of these physical core interactions, 330 were identified only in our proximal interactome, highlighting the benefit of this SARS-CoV-2-host map.

Functional novelties contributed by our SARS-CoV-2-host proximal interactomics

Gene Ontology (GO) and pathways enrichment analyses of the proximal interactome of each viral bait unravel their cell space and point to targeted processes or pathways (**supp table 2**, **supp table 4**, **Fig.S5**). It confirms, *e.g.* the prominent targeting of the vesicular compartments reported in previous interactomics studies and expected given the reliance of SARS-CoV-2 replication and transcription on ER-derived double-membrane vesicle formation. We focused on the novelties provided by our dataset from the GO enrichment analysis. We thus compared the GO terms associated with Biological Processes (BP), Cellular Components (CC) and Pathways (KEGG) emerging specifically in our dataset against the combination of GO terms from all the other interactomic datasets (**supp table 5**). It resulted in the identification of 33 novel pathways overrepresented in the vicinity of viral polypeptides, a number of which relate to reported SARS-CoV-2 infection traits, such as cilium assembly[15], RNA-related processes and various lipids biosynthesis pathways[16] (**supp table 5**). These data thus provide a shortlist of novel PPIs and host factors assigned to pathways known to be altered upon SARS-CoV-2 infection.

To further characterize the links between functionally relevant host proteins within the proximal landscape of viral polypeptides, we intersected our BioID dataset with SARS-CoV-2-related CRISPR-Cas9 studies available at BioGRID ORCS https://orcs.thebiogrid.org/. It appears that 187 host factors required for SARS-CoV-2-induced

cytotoxicity in ≥3 CRISPR-Cas9 screens are present in our BioID dataset, connected to 27 viral bait proteins through 694 proximal interactions (**supp table 2**). The distribution of interactions involving essential host factors is as follows: (**i**) 105 PPIs are laying both in the physical and proximal cores connecting 52 essential host factors to 18 baits (**Fig. 1B**). Of note, this set of PPIs represent super-high confidence PPIs, *i.e.* detected 3+ times by at least two orthogonal methods; (**ii**) 41 PPIs are in the physical core only and connect 16 essential host factors to 16 viral proteins (**Fig.1C**); (**iii**) 297 PPIs are in the proximal core only, connecting 91 essential host factors with to 20 viral proteins (**Fig.1D**); and (**iv**), 261 PPIs were uniquely detected in our interactome, connecting 129 essential host factors to 27 viral proteins (**Fig.1E**). Note that among the 187 essential host factors, several are involved in both the physical, the proximal cores through different PPIs they engage with viral proteins.

Insights in the SARS-CoV-2 proteome hijacking upon anti-viral cellular response

Antiviral cellular response induced by the poly(I:C) treatment expectedly remodeled the proximal landscape of viral proteins with ~10% to ~95% of the interactors being impacted, depending on the viral protein (**supp table 6**). Three distinct remodeling patterns could be distinguished in response to infection-mimicking treatment, excluding the viral proteins not reacting (less than 20% of differential hits) or those with less than 30 hits total: (**i**) net enrichment for S, N, and NSP2; (**ii**) net depletion for NSP10, Orf3a, Orf6, NSP6, NSP7, Orf9b, Orf3d, E and M; and (**iii**), exchange for NSP4, NSP12 and Orf9c (**Fig. 2A and supp table 6**). S and N proteins enter the host cells with the incoming viral particles, and both gained proximal proteins in an antiviral cellular state. S mostly gained ER proteins (93/102), which is a primary target of the virus. The N neighborhood gained the RISC-loading complex components (DICER1, TARBP2, PRKRA), as well as proteins involved in the interferon signaling pathway (IFIT1, IFIT2, IFIT3, EIF2AK2 and TRIM26). Additional novel links between viral factors and major antiviral response pathway components emerged. We detected the recruitment of ISG20 and TOMM70 into close proximity of Orf9b, which is in accordance with its role in the inhibition of IFN-I response to infection[17]. Additionally, we detected ZAP (ZC3HAV1) and TARB2 near N; HERC5 and STAT1 near NSP2; and PARP9 near NSP16 (**supp table 6**).

A comprehensive three-dimensional SARS-CoV-2 interactome

Since BioID bait proteins label proximal partners within a spherical volume with a ~10-20 nm radius[18], proximal interactomics data capture information about the vicinity of the host factor in the physical space of living cells. Based on these properties, we developed a methodology to approximate the relative positions of viral and host proteins inside a theoretical cellular space (see Methods). By modeling the pairwise BioID proximal interactions into a weighted graph structure and then applying a force-directed graph layout algorithm in three dimensions,

we are able to arrive at a set of 3D coordinates for the graph's vertices, representing the positions of viral and host proteins. We generally consider these positions as indicative of the centers of extended cellular regions where molecules of each protein reside. The set of coordinates allows us to infer relative distances between nodes within our abstract 3D space, compare proximities and observe general regions of viral interaction. While difficult to establish a precise, geometric correspondence between our generated 3D volume and the cell's physical volume due to experimental as well as graph modeling limitations (such as in the case of multiple protein localizations which cannot be accurately captured by a single node location), we consider our approximation of the spatial organization of viral and host proteins as forming a "sketch" of their interplay inside the actual physical cell volume.

The global shape of the 3D virus-host network reveals a cluster of 13 viral proteins (Orf3a, Orf3d, M, S, E, NSP3, NSP4, NSP6, Orf8, Orf7a, Orf7b, Orf6 and Orf9c) gathered with human proteins, surrounded by 14 viral proteins dragged out by different sets of partners (**Fig.2B**). From the density of nodes in the 3D map, we thus inferred that removing the host proteins connected to eight or more viral proteins would avoid potential compartment-dependent proximal background. The overall density of the host proteins in the 3D space shows a sharp decrease with radial distance from the cluster's centroid (**Fig.S6**). When restricting to proteins with a maximum of seven viral interactions, we observe a density peak and a smooth decrease with radial distance from the centroid (**Fig.S6**). We use the radius of the density inflection point to define a threshold for the cluster size or dense core. This enables us to identify a list of 790 host factors residing within the dense core boundary, which are in their vast majority assigned to membrane categories (325 to the ER membrane, 165 to the vesicle membrane, 109 to the Golgi membrane and 109 to the plasma membrane region) (**supp table 7**). These membrane-rich organelles are the known primary location of nine viral proteins (Orf3a, M, S, E, NSP3, NSP4, NSP6, Orf7a and Orf7b) [1,2], all of which are included in this core. Expectedly, when removing host proteins with degree 8 or higher, the dense core was relaxed while keeping membrane-bound host proximal partners of a specific set of viral proteins (**Fig.2B**). Hence, this operation facilitates the extraction of discrete proximity-based hypotheses for each viral protein. The resulting 3D virus-host interaction map is available in an interactive web interface (temporarily accessible at http://62.217.122.54/, to be deployed at www.sars-cov-2-interactome.org upon publication) to visualize and explore the SARS-CoV-2-host proteins spatial organization. The regions of pathways-filtered subnetworks can also be captured, revealing their positioning relative to the SARS-CoV-2 proteins, thus highlighting coordination between viral proteins to exploit host cells (**Fig. 2C**). This is well exemplified for protein translation, spatially concentrated near NSP1, NPS7, NSP10 and Orf8, with the translation initiation complex restricted to the vicinity of NSP1, and ribosomal subunits close to Orf8, thereby

distinguishing discrete virus-host interplays regulating translation. In addition, SNARE proteins involved in membrane fusion are specifically near the cluster of viral membrane factors. Other processes are distributed throughout the virus-host proximal network of interactions, such as the cell junction or viral processes.

Assembling data from previous studies in the literature and combining it with our BioID dataset enabled us to get a full picture of the SARS-CoV-2-host interplay identified to date by utilizing the same modeling and 3D visualization approach. The overall network shape determined by our dataset is preserved, indicating that the graph layout approach provides consistent viral-host neighborhoods from all SARS-CoV-2 interactomics studies. Each released interactome is separately accessible in this global SARS-CoV-2-host 3D map. The expected underrepresentation of membrane-associated PPIs obtained from the AP-MS dataset compared to proximal interactomes emerges in the visualization ("Global comparative view" tab at http://62.217.122.54). At a first glance, the membrane-rich central region of the network is equally enriched in CRISPR-Cas9 hits than regions enriched in soluble factors, stressing the importance of those host proteins for SARS-CoV-2 infection. GO-term based filtering of SARS-CoV-2 targeted host factors provides a rather comprehensive picture of the spatial organization of the SARS-CoV-2 proteins in relation to specific parts of the cellular machinery, thus strongly increasing the interpretability of the identified SARS-CoV-2 host interactions. The 3D representation of SARS-CoV-2-host proximal interactions is a pioneering step forward toward a more realistic picture of SARS-CoV-2 invasion patterns in cells, and PPIs network representation in general.

Exploring physical contacts of the viral proteins in the proximity interactome dataset

Combining orthogonal PPI detection methods is a well-established option to increase the accuracy of large-scale PPIs maps. We complemented our BioID proximal interactomics dataset with a split-nanoluciferase assay, which senses binary PPIs by protein-fragment complementation of the nanoluciferase enzyme in human cells (mN2H), and offers excellent detection performance[8]. PPIs screens by mN2H are supervised, systematic (matrix-based) and semi-quantitative (luciferase-based intensity map interactions), providing a strong orthogonality to the BioID (see methods). We sampled two sets of factors from our BioID dataset for mN2H-based PPIs profiling with the SARS-CoV-2 proteins (**Fig. 3A**) : (**i**) 144 host factors highly enriched with less than 4 viral bait proteins (based on log2 fold change over background and q-value significance); and (**ii**) 92 host factors scoring positive in 3+ CRISPR-Cas9 screens. The selected host factors were tested against all 27 viral proteins by mN2H, resulting in comparative virus-host interaction profiles among the selected proximal factors. The full mN2H matrix interrogated about 6372 viral-host protein pairs for direct interaction (**supp tables 8 and 9**), of which 587 correspond to proximal protein pairs as identified by BioID (**supp table 9**).

Among the tested viral-host protein pairs, regardless of their status in BioID, 6% (414/6372) showed direct interactions, involving 118 host proteins. In the protein pairs detected by BioID 21% (123/587) were direct, indicating a 3.5-fold enrichment (**supp table 9**). The mN2H SARS-CoV-2-host contactome determined in this study is provided in **Fig.3B**. It significantly extends our recently released SARS-CoV-2-human contactomes[9,10], offering a wealth of novel SARS-CoV-2-host contacts. Significantly, the mN2H-based systematic PPIs profiling uniquely identified 151 direct interactions involving 41 host factors essential for SARS-CoV-2 lethality in 3+ CRISPR-Cas9 screens, of high therapeutic interest.

Signal intensities detected in split-nanoluciferase assays correlate with the interaction strength[19], allowing qualification of interactions according to their distance from the positive PPIs threshold. From the comparative intensity map, distinct sets of interactions easily emerged, such as NSP1 and NSP7 proteins targeting three subunits of the EIF3 translation complex (EIF3L, EIF3G, EIF3H), or the binding of mitochondrial factors by Orf9b (**Fig.4A**). Other host proteins critical for infection (based on CRISPR hits) are primarily contacted by a single viral protein (**supp table 9**, **Fig.4A**). The intensity map of SARS-CoV-2-host interactions affords an advanced level of precision of the SARS-CoV-2-host interplay with a subset of the human proteome, including proteins essential for virus-induced cytotoxicity.

Hierarchical clustering of mN2H PPIs profiles of SARS-CoV-2 proteins essentially reflects the 3D map (**Fig.4B**), suggesting that direct contacts shape the virus-host spatial organization. It also reveals a specialized manipulation of the host machinery by neighboring viral proteins in the 3D map. PPI profile-based clustering separates viral proteins into two main groups: the first one consists of most of the non-structural proteins that each exhibit few contacts with host proteins, while the second cluster consists of membrane-associated viral proteins that engage in the majority of direct interactions with the host. The most striking host cluster corresponds to proteins involved in ER organization, which is co-opted by all membrane-associated viral factors (**Fig.4B**). Other host clusters appear specific to a given viral protein, such as proteins involved in translation initiation (4 members of the EIF family) or in phosphatidylinositol processes (FLII and WIPI1) that are especially associated with NSP1. NSP3 clusters with proteins involved in negative regulation of proteins' metabolic process (SENP1, RTN4, FXR1 and FXR2). Orf3d targeted proteins involved in intramembrane lipid transporter activity (ATP9A, PLSCR1), while Orf9b binds to many mitochondrial proteins (BCL2L13, MFF, MRPL34, NDUFB8 and TMEM242).

Intraviral interactions profiling

In the course of the virus life cycle, viral proteins cooperate to orchestrate the reprogramming of the cell

machinery. To provide a picture of viral proteins cooperation and expand on previous yeast two-hybrid based datasets[9,10], we explored the intrinsic capacities of the SARS-CoV-2 proteins to interact with each other using mN2H-based PPI profiling (**supp table 10**, **Fig.4C**). A number of viral heterodimers were detected, including the well-characterized NSP16-NSP10[20,21], N-NSP3[22]. Of note, no binary interactions between the core replicase complex proteins (NSP7-NSP8-NSP12) were detected, possibly due to inappropriate the conformation of the proteins when tested in a pairwise manner, as these proteins are part of a more complex structure needing all three proteins to achieve the hexameric architecture of this complex[23]. A hub of intra-viral PPIs emerges within the set of organelles-associated viral factors lying close together in the 3D network and the hierarchical clusters (NSP3, NSP4 and NSP6, E, S, M, Orf7a and Orf7b) (**Fig.4D**). Thus, direct intra-virus interactions are likely building cooperative anchorages within these cellular regions upon infection**.** The intraviral PPI data thus helps to refine the SARS-CoV-2-host interactome. Our SARS-CoV-2-host interactions resource as such provides multi-level insights into the SARS-CoV-2 cellular invasion pattern, both though the spatial positioning of viral proteins relative to host proteome and the direct virus-host and virus-virus contacts underlying such organization during the course of the virus life cycle.

Predictive value of the BioID-based 3D interaction map

When plugged into the BioID data-driven 3D map, the mN2H positive host proteins were closely confined around their viral partners (despite the fact that they were not used for the 3D layout modeling) (**Fig.5A**). We examined the distances between the viral proteins and host factors based on four categories: (**i**) non-interacting, non-proximal proteins (PPIs not detected by either method); (**ii**) proximal proteins (PPIs detected by BioID only) (**iii**) direct-contact proteins within the BioID dataset (PPIs detected by both BioID and mN2H) (**iv**) *de-novo* identified direct-contact proteins (PPIs detected by mN2H only). We report that directly interacting proteins detected by mN2H and identified within the 6372 viral-host protein pairs tested in mN2H were placed at shorter distances on average than non-interacting proteins (based on mN2H negative signal). This remained true when restricting to the subset of proteins detected by BioID, as well as not detected by BioID: mN2H-positive interactions showed an average distance lower than mN2H-negative ones in both cases (**Fig.5B**, **Fig.S6E-F**). While the mN2H-tested matrix showed an average viral-host protein distance of 0.58 in 3D model coordinates, pairs identified by mN2H only, BioID only and mN2H & BioID were at average distances of 0.45, 0.29 and 0.25, respectively. These results support the predictive power of the 3D network layout in foretelling volumes enriched in direct PPIs.

Identifying the interaction interface of host-virus complexes provides an important asset to developing rationalized PPIs-disrupting compounds. We assessed the power of our mN2H matrix of direct interactions to provide virus-host interaction interfaces by *in silico* prediction of the structure of virus-host complexes using

AlphaFold-Multimer[11]. On a selection of 70 virus-host PPIs scoring positive by mN2H and using pDockQ scoring[24,25], 10/70 mN2H-positive pairs scored positive (pDockQ >0.23) (**supp table 11**), providing as many high-confidence interfaces (**Fig. S7**) which can serve as a basis for the development of PPI targeting compounds.

Significance of the SARS-CoV-2-host contacts to lead to the identification of therapeutic targets

We also applied a search for structural homology between virus-host binary interacting pairs (i.e. pairs that scored positive in mN2H) and human complexes with known structures. The previously reported and found in this study interaction between NSP13 and the human deubiquitinase USP13[26], could be derived by homology-based structural modeling with the human UBP14/PRS7 complex (**Fig.5C**). To estimate the potential of targeting NSP13/USP13 with PPIs-disruptive compounds, the involvement of USP13 in SARS-CoV-2 infection has to be demonstrated, given it has not been detected in CRISPR screens.

To address this and meet the need for high-throughput assays that allow screening for compounds with antiviral activity, we developed a time- and cost-effective quantitative readout of SARS-CoV-2-infection in cell lines, which relies on the complementation of the split-nanoluciferase enzyme upon fusion between infected cells, a hallmark of SARS-CoV-2-induced cytopathogenic effect[27–29]. The assay relies on the auto-assembly of two peptides of the nanoluciferase enzyme (a large N-terminal fragment NanoLg and a small C terminal peptide NanoSm, **see method**). Briefly, the assay is based on infection by SARS-CoV-2 of a co-culture of two SARS-CoV-2-sensitive cell lines, each stably expressing one the complementary fragment (VeroE6-NanoLg or VeroE6-NanoSm, respectively). Infection by SARS-CoV-2 induces the formation of syncytia, where complementary nanoluciferase fragments assemble to form an active enzyme (**Fig.5D**). A robust and dose-dependent nanoLuciferase signal was detected upon infection with serial dilutions of the strain Wuhan of SARS-CoV-2 (hCoV-19/France/GES-1973/2020, EPI_ISL_414631) of the co-culture, but not in separated monoculture of each cell line, attesting the performance of this assay (**Fig.S8**). Using this fusion-induced nanoluciferase complementation assay (Nano-Fuse) in VeroE6 cells, we tested the effect of the spautin-1, a previously reported USP13 inhibitor[26]. We observed a dose-dependent decrease of SARS-CoV-2 infection upon spautin-1 treatment (**Fig. 5E**). The widely used protease inhibitor GC-376, assessed in the same assay had a 50% inhibitory concentration (IC50) of 2.3 µM, in accordance with previous reports[30] which ascertains that the nanoluciferase complementation assay properly reflects an active SARS-CoV-2 infection (**Fig.5E**). Spautin-1 inhibited SARS-CoV-2 infection more efficiently than GC376, with a calculated IC50 of ~0.8 µM, while no toxicity of the compound was observed (**Fig. S8**). The dose-response profile suggests that USP13 is important but non-essential for SARS-CoV-2 replication since the inhibition plateaued at 80%, which aligns with its lack of detection in the CRISPR- screens. Nevertheless, our data confirmed the involvement of USP13 in infection by SARS-CoV-2, thus deserving further

screening of PPI disruptive compounds. We have also developed a nano-luciferase complementation assay to detect SARS-CoV-2, whose sensitivity and simplicity enable high-throughput screening of SARS-CoV-2 interfering compounds.

Discussion

The wealth of SARS-CoV-2 interactomics data sketches the complexity of any virus-host interplay. Our data improves the current knowledge of cell hijacking by SARS-CoV-2 first by providing a high coverage of the virus-host landscape, merging data from 27 SARS-CoV-2 polypeptides in basal and infection-mimicking conditions. Combined with the other SARS-CoV-2 interactomes, our dataset significantly increases the number of SARS-CoV-2-host links validated by ≥ 2 orthogonal methods (physical core of virus-host PPI), while overlap with only proximity-dependent interactomes defines a validated proximal core. Our study also identifies 5,114 novel high-confidence proximal interactions. From our BioID dataset, we generated an innovative representation, allowing visualization of the relative spatial distribution of virus and host proteins. BioID identifies proteins within a 10-20 nm spherical radius around a bait protein, which can be leveraged to position baits and interactors relative to each other. The occupation volumes near viral baits can thus be depicted, providing a geometric map of SARS-CoV-2 proteins in the cell. It represents a conceptual advance towards a more realistic description of the virus invasion pattern within the cell's volume and possible strategies to counteract it. Combining the results of all SARS-CoV-2-host interactomes generated to date paves the way for the perturbation of regions emerging as central in cell hijacking, thus offering an avenue for advanced therapeutic paths. In the longer term, the customizable online dashboard will achieve a basal 3D cellular repository obtained by incrementing all human proximal interactomic studies to this initial framework, expanding the scope and reach of this approach. While representing the proteins as fixed points in space presents a limitation since it does not adequately capture protein movement inside the cell or multiple localizations, our representation can still serve to produce indicative centroids of volumes where proteins could be localized and a first attempt to depict this has been implemented as an option in the online dashboard. A more complete modeling would probably entail abandoning the graph structure and assigning 3D probability densities instead of 3D points to each protein, something akin to a probability density "cloud". These densities could then have multiple peaks representing distinct localizations. Using such an extended location formalism is a promising direction for exploiting the wealth of BioID data for a more accurate and realistic cell volume depiction, which we have left for future work.

Beyond the big picture, our work is unique in that it systematically captures the direct virus-host contacts within a subset of the proximal interactome, identifying 414 direct virus-host PPIs. These direct PPIs are characterized by shorter than average edges in the geometric SARS-CoV-2-host map, supporting the predictive power of the

map on direct contact probability as forebode[31].

Among the 235 cellular proteins sampled from BioID dataset, 118 displayed direct contacts, of which 41 are essential high confidence factors for SARS-CoV-2-induced toxicity previously identified in 3+ CRISPR-Cas9 screens. The two only recently released contactome maps[9,10], based on yeast two-hybrid, remarkably showed that the direct targets of viral proteins are part of network communities enriched in proteins genetically associated with the risk of developing severe COVID-19[14], and that such direct SARS-CoV-2 targets are enriched in gene products differentially expressed in samples from COVID-19 patients[15], respectively. Importantly, our pipeline was thus the only one leading to the capture of direct contacts between SARS-CoV-2 proteins and human proteins essential for infection, those PPIs being attractive targets for anti-viral therapeutics exploration. We believe it is likely due to the ability of our pipeline to identify viral proximal factors by BioID independently of their subcellular localization and solubility. Indeed, of the 151 direct PPIs engaging the 41 essential host proteins, 91 (60%) involve SARS-CoV-2 proteins associated with membrane-associated sub-cellular compartments (NSP3, NSP4, NSP6, S, E, M, Orf3a, Orf7a and Orf7b), and 31 of 41 (76%) targeted essential host proteins are associated with intracellular organelles.

Profiling by mN2H revealed a clustering of the viral proteins with the relative strength of the interactions highlighting the strongest ones, thereby helping to uncover the main drivers of SARS-CoV-2 mediated cell hijacking. This is exemplified by the set of transmembrane proteins (TMED2, TMEM41B, TMEM106B, TMEM242, TMEM33 and TMEM159) exhibiting distinct interaction patterns with viral proteins. The TMEM41B, a poorly characterized ER-localized protein, prominently binds to Orf9b and Orf3a and has been shown to be essential for infection by SARS-CoV-2, and by other human coronaviruses (HCoV-OC43, HCoV-NL63, and HCoV-229E [32]) as well as by some flaviviruses[33]. It represents a cellular regulation hub co-opted by multiple viruses that is worth countering. Weak interactions represent low affinity or labile PPIs, which may also be functionally relevant. Thus, prioritizing PPIs to be further investigated should consider both the interaction profiles (interaction strength and redundancy) and host protein essentiality for infection.

Knowledge of the structure of SARS-CoV-2-host complexes would be a clear asset for screening PPIs-disruptive compounds. The mN2H dataset provides a resource for PPIs modeling, which we trained using a selection of N2H-identified virus-host complexes. This analysis identified 10 virus-host complexes of strong existence probability (DockQ>0.23) and provided structure models. The AlphaFold2-predicted structure of those proteins that are structurally known are close to their experimentally defined structures. Examples include the Rab6A and Rab7A proteins structured in complexes with NSP4 of SARS-CoV-2, which are perfectly superimposable with their experimentally determined crystal structures.

The structure of virus-host protein complexes can also be predicted by homology-based modeling, when homology is found with structurally characterized co-complexes. This was the case for NSP13/USP13 interaction by structure-based homology with the human UBP14/PRS7 complex. To assess the role of USP13 in SARS-CoV-2 infection, we developed a quick, cost-effective and sensitive assay for measuring SARS-CoV-2 infection, based on the complementation of the nanoluciferase enzyme mediated by cell fusion following infection. We then characterized USP13 as important for SARS-CoV-2 replication since its inhibition by Spautin-1 drastically impaired SARS-CoV-2 replication with an IC50 of ~0.8 µM. These data corroborate a previous study[26] and point to Spautin-1 as a potential adjuvant therapy. Together, these results validate the NSP13/USP13 interaction as relevant for structure-driven screening of PPIs disruptive drugs. Importantly, they provide proof of principle that our proteomic BioID/mN2H pipeline represents a valuable resource for selecting relevant PPIs for antiviral compound screening. This pipeline is easily adaptable to all emerging human pathogens to prompt the identification of therapeutic targets and, therefore, should account for improved preparedness.

Literature cited

1. Zhang, J., Cruz-Cosme, R., Zhuang, M.-W., Liu, D., Liu, Y., Teng, S., Wang, P.-H., and Tang, Q. (2020). A systemic and molecular study of subcellular localization of SARS-CoV-2 proteins. Signal Transduct Target Ther *5*, 269. 10.1038/s41392-020-00372-8.

2. Mariano, G., Farthing, R.J., Lale-Farjat, S.L.M., and Bergeron, J.R.C. (2020). Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. Front Mol Biosci *7*, 605236. 10.3389/fmolb.2020.605236.

3. Stukalov, A., Girault, V., Grass, V., Karayel, O., Bergant, V., Urban, C., Haas, D.A., Huang, Y., Oubraham, L., Wang, A., et al. (2021). Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. Nature *594*, 246–252. 10.1038/s41586-021-03493-4.

4. Li, J., Guo, M., Tian, X., Wang, X., Yang, X., Wu, P., Liu, C., Xiao, Z., Qu, Y., Yin, Y., et al. (2021). Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis. Med (N Y) *2*, 99–112.e7. 10.1016/j.medj.2020.07.002.

5. Gordon, D.E., Hiatt, J., Bouhaddou, M., Rezelj, V.V., Ulferts, S., Braberg, H., Jureka, A.S., Obernier, K., Guo, J.Z., Batra, J., et al. (2020). Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science *370*. 10.1126/science.abe9403.

6. Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459–468. 10.1038/s41586-020-2286-9.

7. Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. *196*, 801–810. 10.1083/jcb.201112098.

8. Choi, S.G., Olivet, J., Cassonnet, P., Vidalain, P.-O., Luck, K., Lambourne, L., Spirohn, K., Lemmens, I., Dos Santos, M., Demeret, C., et al. (2019). Maximizing binary interactome mapping with a minimal number of assays. Nat. Commun. *10*, 3907. 10.1038/s41467-019-11809-2.

9. Kim, D.-K., Weller, B., Lin, C.-W., Sheykhkarimli, D., Knapp, J.J., Dugied, G., Zanzoni, A., Pons, C., Tofaute, M.J., Maseko, S.B., et al. (2022). A proteome-scale map of the SARS-CoV-2–human contactome. Nat. Biotechnol., 1–10. 10.1038/s41587-022-01475-z.

10. Zhou, Y., Liu, Y., Gupta, S., Paramo, M.I., Hou, Y., Mao, C., Luo, Y., Judd, J., Wierbowski, S., Bertolotti, M., et al. (2023). A comprehensive SARS-CoV-2-human protein-protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets. Nat. Biotechnol. *41*, 128–139. 10.1038/s41587-022-01474-0.

11. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer. BioRxiv, 2021.10.04.463034. 10.1101/2021.10.04.463034.

12. St-Germain, J.R., Astori, A., Samavarchi-Tehrani, P., Abdouni, H., Macwan, V., Kim, D.-K., Knapp, J.J., Roth, F.P., Gingras, A.-C., and Raught, B. (2020). A SARS-CoV-2 BioID-based virus-host membrane protein interactome and virus peptide compendium: new proteomics resources for COVID-19 research. BioRxiv, 2020.08.28.269175. 10.1101/2020.08.28.269175.

13. Samavarchi-Tehrani, P., Abdouni, H., Knight, J.D.R., Astori, A., Samson, R., Lin, Z.-Y., Kim, D.-K., Knapp, J.J., St-Germain, J., Go, C.D., et al. (2020). A SARS-CoV-2 – host proximity interactome. BioRxiv. 10.1101/2020.09.03.282103.

14. May, D.G., Martin-Sancho, L., Anschau, V., Liu, S., Chrisopulos, R.J., Scott, K.L., Halfmann, C.T., Díaz Peña, R., Pratt, D., Campos, A.R., et al. (2022). A BioID-Derived Proximity Interactome for SARS-CoV-2 Proteins. Viruses *14*, 611. https://doi.org/10.3390/v14030611.

15. Robinot, R., Hubert, M., de Melo, G.D., Lazarini, F., Bruel, T., Smith, N., Levallois, S., Larrous, F., Fernandes, J., Gellenoncourt, S., et al. (2021). SARS-CoV-2 infection induces the dedifferentiation of multiciliated cells and impairs mucociliary clearance. Nat. Commun. *12*, 4354. 10.1038/s41467-021-24521-x.

16. Casari, I., Manfredi, M., Metharom, P., and Falasca, M. (2021). Dissecting lipid metabolism alterations in SARS-CoV-2. Prog. Lipid Res. *82*, 101092. 10.1016/j.plipres.2021.101092.

17. Jiang, H.-W., Zhang, H.-N., Meng, Q.-F., Xie, J., Li, Y., Chen, H., Zheng, Y.-X., Wang, X.-N., Qi, H., Zhang, J., et al. (2020). SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. Cell. Mol. Immunol. *17*, 998–1000. 10.1038/s41423-020-0514-8.

18. Kim, D.I., Birendra, K.C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K.J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc. Natl. Acad. Sci. U. S. A. *111*, E2453–E2461. 10.1073/pnas.1406459111.

19. Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., Turchetto, J., Iv, F., Ricquier, K., Straub, M.-L., et al. (2015). Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. Nat. Methods *12*, 787–793. 10.1038/nmeth.3438.

20. Krafcikova, P., Silhan, J., Nencka, R., and Boura, E. (2020). Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. Nat. Commun. *11*, 3717. 10.1038/s41467-020-17495-9.

21. Viswanathan, T., Arya, S., Chan, S.-H., Qi, S., Dai, N., Misra, A., Park, J.-G., Oladunni, F., Kovalskyy, D., Hromas, R.A., et al. (2020). Structural basis of RNA cap modification by SARS-CoV-2. Nat. Commun. *11*, 3718. 10.1038/s41467-020-17496-8.

22. Ni, X., Han, Y., Zhou, R., Zhou, Y., and Lei, J. (2023). Structural insights into ribonucleoprotein dissociation by nucleocapsid protein interacting with non-structural protein 3 in SARS-CoV-2. Commun Biol *6*, 193. 10.1038/s42003-023-04570-2.

23. 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. 10.1038/s41580-021-00432-z.

24. Basu, S., and Wallner, B. (2016). DockQ: A Quality Measure for Protein-Protein Docking Models. PLoS One *11*, e0161879. 10.1371/journal.pone.0161879.

25. Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. Nat. Commun. *13*, 1265. 10.1038/s41467-022-28865-w.

26. Guo, G., Gao, M., Gao, X., Zhu, B., Huang, J., Luo, K., Zhang, Y., Sun, J., Deng, M., and Lou, Z. (2021). SARS-CoV-2 non-structural protein 13 (nsp13) hijacks host deubiquitinase USP13 and counteracts host antiviral immune response. Signal Transduct Target Ther *6*, 119. 10.1038/s41392-021-00509-3.

27. Buchrieser, J., Dufloo, J., Hubert, M., Monel, B., Planas, D., Rajah, M.M., Planchais, C., Porrot, F., Guivel-Benhassine, F., Van der Werf, S., et al. (2020). Syncytia formation by SARS-CoV-2-infected cells. EMBO J. *39*, e106267. 10.15252/embj.2020106267.

28. Bussani, R., Schneider, E., Zentilin, L., Collesi, C., Ali, H., Braga, L., Volpe, M.C., Colliva, A., Zanconati, F., Berlot, G., et al. (2020). Persistence of viral RNA, pneumocyte syncytia and thrombosis are hallmarks of advanced COVID-19 pathology. EBioMedicine *61*, 103104. 10.1016/j.ebiom.2020.103104.

29. Braga, L., Ali, H., Secco, I., Chiavacci, E., Neves, G., Goldhill, D., Penn, R., Jimenez-Guardeño, J.M., Ortega-Prieto, A.M., Bussani, R., et al. (2021). Drugs that inhibit TMEM16 proteins block SARS-CoV-2 spike-induced syncytia. Nature *594*, 88–93. 10.1038/s41586-021-03491-6.

30. Ma, C., Sacco, M.D., Hurst, B., Townsend, J.A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., Marty, M.T., Chen, Y., et al. (2020). Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. Cell Res. *30*, 678–692. 10.1038/s41422-020-0356-z.

31. Gupta, G.D., Coyaud, É., Gonçalves, J., Mojarad, B.A., Liu, Y., Wu, Q., Gheiratmand, L., Comartin, D., Tkach, J.M., Cheung, S.W.T., et al. (2015). A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface. Cell *163*, 1484–1499. 10.1016/j.cell.2015.10.065.

32. Schneider, W.M., Luna, J.M., Hoffmann, H.-H., Sánchez-Rivera, F.J., Leal, A.A., Ashbrook, A.W., Le Pen, J., Ricardo-Lax, I., Michailidis, E., Peace, A., et al. (2021). Genome-Scale Identification of SARS-CoV-2 and Pan-coronavirus Host Factor Networks. Cell *184*, 120–132.e14. 10.1016/j.cell.2020.12.006.

33. Hoffmann, H.-H., Schneider, W.M., Rozen-Gagnon, K., Miles, L.A., Schuster, F., Razooky, B., Jacobson, E., Wu, X., Yi, S., Rudin, C.M., et al. (2021). TMEM41B Is a Pan-flavivirus Host Factor. Cell *184*, 133–148.e20. 10.1016/j.cell.2020.12.005.

34. Kim, D.-K., Knapp, J.J., Kuang, D., Chawla, A., Cassonnet, P., Lee, H., Sheykhkarimli, D., Samavarchi-Tehrani, P., Abdouni, H., Rayhan, A., et al. (2020). A Comprehensive, Flexible Collection of SARS-CoV-2 Coding Regions. G3 *10*, 3399–3402. 10.1534/g3.120.401554.

35. Vabret, N., Najburg, V., Solovyov, A., Šulc, P., Balan, S., Beauclair, G., Chazal, M., Varet, H., Legendre, R., Sismeiro, O., et al. (2019). Y-RNAs Lead an Endogenous Program of RIG-I Agonism Mobilized upon RNA Virus Infection and Targeted by HIV. BioRxiv, 773820. 10.1101/773820.

36. Coyaud, E., Ranadheera, C., Cheng, D., Gonçalves, J., Dyakov, B.J.A., Laurent, E.M.N., St-Germain, J., Pelletier, L., Gingras, A.-C., Brumell, J.H., et al. (2018). Global Interactomics Uncovers Extensive Organellar Targeting by Zika Virus. Mol. Cell. Proteomics *17*, 2242–2255. 10.1074/mcp.TIR118.000800.

37. 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. 10.1093/nar/gkab1038.

38. Hu, Y. (2005). Efficient, high-quality force-directed graph drawing. Mathematica journal *10*, 37–71.

39. Walshaw, C. (2001). A Multilevel Algorithm for Force-Directed Graph Drawing. In Graph Drawing (Springer Berlin Heidelberg), pp. 171–182. 10.1007/3-540-44541-2_17.

40. Barnes, J., and Hut, P. (1986). A hierarchical O(N log N) force-calculation algorithm. Nature *324*, 446–449. 10.1038/324446a0.

41. Fruchterman, T.M.J., and Reingold, E.M. (1991). Graph drawing by force-directed placement. Softw. Pract. Exp. *21*, 1129–1164. 10.1002/spe.4380211102.

42. Van Rossum, G. Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.

43. Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX (Los Alamos National Lab.(LANL), Los Alamos, NM (United States)).

44. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362. 10.1038/s41586-020-2649-2.

45. Reback, J., jbrockmendel, McKinney, W., Van den Bossche, J., Augspurger, T., Roeschke, M., Hawkins, S., Cloud, P., gfyoung, Sinhrks, et al. (2022). pandas-dev/pandas: Pandas 1.4.2 10.5281/zenodo.6408044.

46. Inc, P.T. (2015). Collaborative data science. Montreal: Plotly Technologies Inc Montral.

47. Gagnot, G., Hervin, V., Coutant, E.P., Goyard, S., Jacob, Y., Rose, T., Hibti, F.E., Quatela, A., and Janin, Y.L. (2021). Core-Modified Coelenterazine Luciferin Analogues: Synthesis and Chemiluminescence Properties. Chemistry *27*, 2112–2123. 10.1002/chem.202004311.

48. Coutant, E.P., Goyard, S., Hervin, V., Gagnot, G., Baatallah, R., Jacob, Y., Rose, T., and Janin, Y.L. (2019). Gram-scale synthesis of luciferins derived from coelenterazine and original insights into their bioluminescence properties. Org. Biomol. Chem. *17*, 3709–3713. 10.1039/c9ob00459a.

49. Coutant, E.P., Gagnot, G., Hervin, V., Baatallah, R., Goyard, S., Jacob, Y., Rose, T., and Janin, Y.L. (2020). Bioluminescence Profiling of NanoKAZ/NanoLuc Luciferase Using a Chemical Library of Coelenterazine Analogues. Chemistry *26*, 948–958. 10.1002/chem.201904844.

50. Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265–269. 10.1038/s41586-020-2008-3.

Methods

Plasmids

SARS-CoV-2 viral proteins coding sequences were cloned by Gateway recombination system (Life Technologies) into pDEST pcDNA5 FRT/TO N-ter and C-ter BirA*Flag vectors for the BioID (Coyaud et al., 2018), and into pciNeo-N2 or pciNeo-C2 [8] using the collection of Gateway-compatible entry clones reported in [34] (see Supp Table 1 for sequences). For N2H, cDNA encoding the human proteins to be tested were available pDONR vectors from the human ORFeome collection v8.1 from the Center for Cancer Systems Biology (CCSB) (ORFeome Collaboration, 2016)) or inhouse made clones. Each hORF was introduced into pciNeo-N1 pDEST plasmids by Gateway reaction. All clones were sequence verified.

Cell lines

HEK293 cells stably were stably transfected by the different BirA*tag- encoding plasmids as in (Coyaud et al. 2018). Briefly, Flp-In™ T-REx™ HEK293 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich), GlutaMAX™ and Penicillin-Streptomycin (1x). Using the Flp-In system (Invitrogen), Flp-In™ T-REx™ HEK293 stably expressing BirA*Flag or FlagBirA* alone (for control samples), or N- and C-terminally tagged viral bait proteins were generated by co-transfecting pOG44 with each pcDNA5 FRT/TO BirA*Flag-viral protein encoding plasmid and selected with 200 µg/ml hygromycin B

Confocal microscopy

Each N-ter tagged viral protein expression was inducted by 1µg/mL tetracycline and 50µM biotin for 24 hrs (in basal or poly(I:C) conditions) in stably transfected Flp-In™ T-REx™ HEK293 cells. Briefly, cells were grown on poly-lysine coated coverslips, rinsed with PBS before fixation with 4% paraformaldehyde for 15 min at RT. After fixation, they were washed 3 times with PBS 0.1% Triton X-100. Cells were then blocked in PBS 0.01% Triton, and 5% bovine serum albumin (BSA) for 30 min at 4°C. Transfected cells were then incubated overnight at 4°C with monoclonal anti-FLAG M2 mouse antibody (dilution 1:2,000, Sigma-Aldrich). After 3 washes for 10 min with PBS 0.01% Triton X-100, cells were incubated for 1 h at 37 °C with the Alexa Fluor 647-conjugated secondary antibody donkey anti-mouse IgG (H+L) (1:2,000, Thermo Fisher Scientific) and streptavidin-Alexa Fluor 488 (1:20,000; Invitrogen). Cells were rinsed with PBS, and nuclei were stained with 1 µg/ml 4',6-diamidino-2-phenylindole DAPI (Thermo Fisher Scientific) in PBS for 5 min. Finally, cells were washed, desalted and mounted using Dako Fluorescent Mounting Medium (Agilent Dako, Santa Clara, CA, USA). Acquisition was performed on

a Zeiss LSM700 confocal microscope connected to a Zeiss Axiovert 200 M equipped with an EC Plan-Neofluar 40×/1.30 numerical aperture and an oil immersion objective (Carl Zeiss AG, Oberkochen, Germany). Processing of the images was performed using Zeiss Zen 2 software and assembled using Adobe Illustrator.

IFN-dependent luciferase activity assays

ISRE reporter HEK293 cells referred to as STING37 (Lucas-Hourani et al., 2013) were plated in a 24-well plate (2.5 x 10^5 per well). After 24h, cells were transfected using Lipofectamine 2000 (11668-019, Invitrogen by Thermo Fisher Scientific) with 100 ng of pCIneo-MAVS [35] or the empty vector (pCIneo) with or without 200 ng of pFlag-BirA plasmids expressing viral proteins. 24 h post-transfection cells were lysed by Passive Lysis buffer (E1941, Promega), and luciferase activity was measured with Bright-Glo Luciferase assay system (E2650, Promega).

Bio-ID sample generation

BioID samples were prepared from Flp-In™ T-REx™ HEK293 expressing each BirA*-flagged SARS-CoV-2 protein or BirA*alone as in [36]. Briefly, three independent replicates of two 150 cm2 plates of sub-confluent (60%) cells were incubated for 24 hrs in complete media supplemented with 1 µg/ml tetracycline (Sigma), 50 µM biotin (Thermo Fisher Scientific). For poly(I:C) treatment conditions, poly(I:C) (Sigma Sigma-Aldrich; P1530) was transfected at 2µg/mL for 5hrs using PolyJet reagent prior to tetracycline and biotin 5 hours before tetracycline and biotin addition. Cells were collected and pelleted (300 x *g*, 3 min), washed twice with PBS, and dried pellets were snap frozen. Each cell pellet was resuspended in 5 ml of lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% SDS, 1:500 protease inhibitor cocktail (Sigma-Aldrich), 1:1,000 Turbonuclease (BPS Bioscience) and incubated on an end-over-end rotator at 4°C for 1 hour, briefly sonicated to disrupt any visible aggregates, then centrifuged at 45,000 x g for 30 min at 4°C. Supernatant was transferred to a fresh 15 mL conical tube. 25 µl of packed, pre-equilibrated Streptavidin Ultralink Resin (Pierce) were added, and the mixture incubated for 3 hours at 4°C with rotation. Beads were pelleted by centrifugation at 300 x *g* for 2 min and transferred with 1 mL of lysis buffer to a fresh eppendorf tube. Beads were washed once with 1 mL of lysis buffer and twice with 1 mL of 50 mM ammonium bicarbonate (pH=8.3), then transferred in ammonium bicarbonate to a fresh centrifuge tube and washed two more times with 1 ml of ammonium bicarbonate buffer. Tryptic digestion was performed by incubating the beads with 1 µg MS-grade TPCK trypsin (Promega, Madison, WI) dissolved in 200 µl of 50 mM ammonium bicarbonate (pH 8.3) overnight at 37°C. The following morning, 0.5 µg MS-grade TPCK trypsin was added to the beads and incubated 2 additional hours at 37°C. Following centrifugation at 2,000 x *g* for 2 min, the supernatant was collected and transferred to a fresh

eppendorf tube. Two additional washes were performed with 150 µL of 50 mM ammonium bicarbonate and pooled with the first eluate. The sample was lyophilized and resuspended in buffer A (2% ACN 0.1% formic acid). 1/3rd of each sample was analyzed per mass spectrometer run.

BioID data acquisition

MS samples were prepared from three biological replicates of each bait protein fused either with an N-terminal or a C-terminal BirA*Flag epitope tag in basal condition or following poly(I:C) (Sigma-Aldrich; P1530) transfection at 2µg/mL for 5hrs using PolyJet reagent (Signagen) prior to tetracycline and biotin induction, and analyzed on a Thermo Q-Exactive mass spectrometer. Samples were separated by online reversed-phase chromatography using a Thermo Scientific Easy-nLC1000 system equipped with a Proxeon trap column (75 µm ID x 2 cm, 3 µm, Thermo Scientific) and a C18 packed-tip column (Acclaim PepMap, 75 µm ID x 50 cm, 2 µm, Thermo Scientific). The digested peptides were separated using an increasing amount of acetonitrile in 0.1% formic acid from 2 to 30% for 2 hours at a 300 nL/min flow rate. A voltage of 2.4 kV was applied by the liquid junction to electrospray the eluent using the nanospray source. A high-resolution mass spectrometer Q-Exactive™ Thermo Scientific™ was coupled to the chromatography system to acquire the 10 most intense ions of MS1 analysis (Top 10) in data-dependent mode. The MS analyses were performed in positive mode at a resolving power of 70,000 FWHM, using an automatic gain control target of 3e6, the default charge state was set at 2 and a maximum injection time at 120 ms. For full scan MS, the scan range was set between m/z 300 to 1600. For ddMS2, the scan range was between m/z 200 to 2000, 1 microscan was acquired at 17,500 FWHM, an AGC was set at 5e4 ions, and an isolation window of m/z 4,0 was used. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[37] partner repository with the dataset identifier PXD033452 (reviewer username: pxd033452@ebi.ac.uk; pw: Z1tHrDTc)

BioID data analysis

The proteins were identified by comparing all MS/MS data with the Homo sapiens proteome database (Uniprot, release March 2020, Canonical+Isoforms, comprising 42,360 entries + viral bait protein sequences added manually), using the MaxQuant software version 1.5.8.3. The digestion parameters were defined using trypsin with 2 maximum missed cleavages. The oxidation of methionine and N-terminal protein acetylation were defined as variable modifications. The Label-free quantification (LFQ) was done keeping the software's default parameters. As for initial mass tolerance, 6 ppm was selected for MS mode, and 20 ppm was set for fragmentation data to match MS/MS tolerance. The identification parameters of the proteins and peptides were performed with a false discovery rate (FDR) at 1%, and a minimum of 2 unique peptides per protein. The LFQ

values from the 30 control runs (regrouping FlagBirA* and BirA*Flag alone samples, from stable and transiently transfected cell lines, and w/wo poly(I:C) pretreatment) were collapsed to the three highest values for each given ID. These three values were defined as the control group for comparison with viral bait proteins triplicates. The statistical analysis was done by Perseus software (version 1.6.2.3). Briefly, the LFQ intensity of each sample was downloaded in Perseus, and the data matrix was filtered by removing the potential contaminants, reverse and only identified by site. The data were then transformed using the log2(x) function. Before statistical analysis, 81 groups (27 bait proteins, N-ter, C-ter and poly(I:C) for each) were defined with 3 replicates per group. Only preys with detected values in all three replicates of a given viral bait protein were kept for further analysis. Missing values were then replaced from the normal distribution separately for each column. Two-sample Student's T-test was then performed comparing all three biological replicates of each bait and condition against the three control runs. High-confidence proximal interactors were defined by permutation-based FDR with a cut-off of 0.01. Perseus output with all experimental values is reported in **Supp Table 2** (tab B).

The matrix (**Supp Table 2**) shows the average log2 fold change against control and the corresponding p- and q-values for each bait and condition. The poly(I:C) data were also compared against the control (green font when significant) and the corresponding condition (N-terminal tag for all viral bait proteins, except for NSP1). The poly(I:C) column indicates the status of each interaction when compared to the basal condition. 'Decreased' and 'Increased' were defined when the log2 fc against the basal condition was <(−1) or >(+1), respectively, and the p-value comparing the 3 replicates of the two conditions below 0.05. 'Unchanged' corresponds to log2fc values between −1 and +1, and empty cells depict interactors undetected in the basal and poly(I:C) condition. The InDegrees column depicts the number of bait proteins detecting a given interactor, regardless of the condition (N-ter, C-ter or poly(I:C)). This criterion was chosen to filter out the most connected interactors (8+), likely to be organelle-specific background not filtered using the BirA* alone control samples. 2D networks presented in the different figures were generated using the Cytoscape software (v.3.9.1; https://cytoscape.org/).

Modeling the virus-host interplay in 3D from the BioID Data set

In order to approximate the regions where proteins mainly reside in the cell volume, we used computational techniques for the layout of graphs in three dimensions. A custom force-directed and multi-stage algorithm was employed due to its general applicability and capability for the generation of high-quality layouts[38,39–41]. Another advantage offered by the class of force-directed algorithms, making it suitable for our specific case, is the inherent physical attributes and the flexibility of tuning provided by the underlying physics-inspired model producing the layout coordinates. In this approach, the observed BioID proximal interactions (when present) modulate the strength of an attractive force between proximal protein pairs, while repulsive forces universal

among all node pairs ensure the separation of regions and the expansion into the available volume. In this way, host proteins previously known to have proximal interactions are pulled close to each other and viral proteins with overlapping sets of host interactors are also brought into relative proximity. Interactors for which there is no pairwise proximal relation are pushed apart, their placement determined by their own set of individual proximal interactions. By simultaneously aggregating all the proximity interactors and their interactions subject to these forces, a global approximate picture of the relative spatial organization of viral proteins with respect to the host proteome is emerging.

Interactome network setup and visualization

The starting point for constructing the interactome network is the BioID high confidence PPI data table, each row of the table containing data of a single bait-target interaction. Aggregating all the PPI data, we obtained a total of 10,125 interactions involving 27 baits and 2,593 targets (total of 2,620 nodes) to which we also appended 290 mN2H positive interactions not detected by BioID (these latter PPIs are not participating in the layout and therefore do not influence the node coordinates, but are used for displaying purposes). Based on this dataset, we construct an undirected graph where interactors and interactions are represented as the graph's vertices (nodes) and edges (links) respectively. In addition to links due to the observed BioID bait-target interactions, we augment the graph with 3,492 edges representing previously known target-target interactions. We thus obtain a final graph with a total of 2,620 vertices and 13,617 edges.

To visualize the graph, our objective is to determine an optimum placement of the nodes in three-dimensional space that would minimize edge crossings and reveal the interactome's structure to the best possible extent. We employ a custom, multi-stage, force-directed graph layout algorithm in 3D, based largely on the spring-electric model outlined by Hu [38]. We modulate the strength of the attractive (spring) forces in the corresponding physical model with a factor *w* given by the combination of the BioID enrichment ratio (or fold change, *fc*) according to the formula: $w = \left[\sum_i log_2(fc)_i\right]^p$, where p=½ and i = {N-ter, C-ter, poly(I:C)}, when present for the interaction. For reported target-target interactions where this information is not available we set this factor equal to 1. We also ignore the aforementioned purely mN2H positive (non-BioID) interactions in order to preserve the orthogonality of the two approaches. The node coordinates in 3D obtained by minimizing the energy of the spring-electric physical model are ingested by a plotting package that generates a 3D visualization of the network, assigning a point to each node and a line connecting a pair of nodes when there is an interaction between the corresponding interactors. Our code was written in Python 3.8 [42] and makes use of several modules,

primarily: NetworkX for graph operations [43], NumPy for numerical computations [44], pandas for data manipulation [45] and Plotly for visualization [46].

Interactive web application

In order to make our results available to the community in a usable and informative manner, we have implemented a web application, accessible at http://www.sars-cov-2-interactome.org/. It contains the network visualization, as well as multiple options for selecting and filtering the dataset to view sub-networks, providing the ability to focus on particular areas of interest and explore various levels of detail. The interactive 3D visualization of the interactome is a main ingredient of the dashboard, as it enables navigation within a browser window, zooming and rotating in space, giving an intuitive illustration of the relative localizations of proteins inside the cell volume. In addition, we integrate information about the previously reported SARS-CoV-2 status of interactors and interactions, as well as previously reported interactions between the targets (cf. **Supp Table 3** for details). The network customization options contain various filters, separated broadly into two levels: on the first level, there are options for: (**i**) focusing on targets of higher specificity by filtering out targets with a total number of bait interactions (or degree) above a configurable threshold (set to 7 by default); (**ii**) restricting to targets based on known relevance for SARS-CoV-2 (based on total number of CRISPR screens the given interactor has been identified as functionally important for SARS-CoV-2 infection); (**iii**) filtering interactions based on their poly(I:C) treatment result; or (**iv**), selecting particular baits of interest and their interactions only. When multiple filters are activated at this level, the resulting dataset is formed by the intersection of the selected criteria. In addition, at this level, we provide options for displaying the previously reported target-target interactions, as well as highlighting previously reported and N2H-validated bait-target interactions. At the second level, there is the functionality to select any number of annotations of interest from three main categories (pathways, biological processes, cellular components) in order to generate an annotated subnetwork, while remaining subject to the filtering selections from the preceding level. In this second level, multiple annotation selections are combined so that the resulting network is the union of the individually annotated subnetworks. Specific targets of interest can also be selected individually, or added to an already generated subnetwork. Notably, there is an option to select any number of annotations for which to display their approximate localization overlaid on the network, either as 3D density volumes, or as labels positioned at the volume's centroids ('label cloud'). At any given time, the interactive visualization reflects the user's filtering selections, offering an intuitive view of the particular subset of data under consideration, while the filtered dataset is also visible in table format below the network and downloadable. Finally, there are options for mostly cosmetic attributes of the network visualization, such as the ability to toggle text labels and edges, present a faded view of the full network in the background (thus giving a better global view of the location of nodes and edges in the interactome), and display 3D spherical shapes for

the interactors (as opposed to flat 2D circular markers), better representing their physical size and actual location in space. A summary panel of the network with a breakdown of the totals for nodes and edges is presented alongside the network, as well as a panel displayed upon clicking on a node or edge with corresponding metadata, including AlphaFold protein structure predictions. A detailed usage tutorial is also provided as part of the website. The dashboard web application has been implemented using the Dash module in Python 3.8.

Mammalian cell-based N2H assay

HEK293T cells were seeded at $6x10^4$ cells per well in 96-well, flat-bottom, cell culture microplates (Greiner Bio-One, #655083), and cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum at 37 °C and 5% CO_2. Twenty-four hours later, cells were transfected with 100 ng of each N1- and N2/C2-expressing plasmids using linear polyethylenimine (PEI MAX 40000; Polysciences Inc; Cat# 24765) to co-express the protein pairs fused with complementary NanoLuc fragments, F1 and F2 [8]. Twenty-four hours after transfection, the culture medium was removed, and 30 µL of 100× diluted NanoLuc substrate[47] was added to each well of a 96-well microplate containing the transfected cells. Luciferase enzymatic activity was measured using a CentroXS luminometer (Berthold; 2 s integration time). The luciferin substrate used (Q-108) in all the bioluminescence experiments was obtained in a concentrated solution, from the corresponding O-acetylated derivative hikarazine-108, following acidic hydrolysis as previously described[48,49]. The binary virus-host PPIs were assessed and were independently performed two times, each made in duplicates to quadruplicates. PPIs were monitored using the cellular proteins (CP) fused the F1 nanoluciferase fragment at their N-terminus (N1-CP), while the viral proteins were fused to the nanoluciferase F2 fused either at N- , their C-terminus or both. Configuration of the virus-host PPIs matrix was either N1N2 or N1C2, and was selected from the viral bait Bir-A* fusion configuration leading to host partners identification in BioID.

To score positive PPIs from the tested matrixes, we applied the following pipeline:

1-Log2-transformed averaged RLU signals were normalized to the mean columns signal and mean row signals (Log2 RLU PPI - mean Log2 column - mean Log2 rows). Column and row averages reflect the overall level of viral and cellular protein interaction signals. A Z-score (relative to the mean and standard deviation of the full Normalized PPIs matrix) was applied on normalized data to obtain normal distribution and to calculate a P-value. A significant P-value (< 0.05) or equivalently a Z-score upper to 1.64 were scored as positive PPIs. In this way, positive PPIs are distinguished over non-interacting pairs with high stringency, leading to the identification of high-confidence PPIs.

Hierarchical clustering of virus-host PPIs was performed using the R software by using euclidean distance

parameter of the dist function. Clustering was then performed by ward.D method with R software by using pheatmap function.

A novel sensitive assay to measure SARS-CoV-2 infection

cDNA encoding for the nanoluciferase fragment 1-158 (NanoLg) and the 13 aa long c-terminal Nanoluciferase peptide (aa 159-171, Nano-Sm) were cloned in the pLVX-puro vector (Clontech) in fusion with GFP11 (at the C-ter of LN) or GFP1-10 (at the N-ter of SN) respectively. Lentiviral vectors were produced by co-transfection of the pLVX-puro NanoLg or pLVX-puro NanoSm plasmids with gag-pol packaging plasmid psPAX2 (Addgene, #12260) and VSV-G envelope expressing plasmid in HEK293T cells using calcium phosphate precipitation. Lentiviral titers were measured by p24 ELISA according to the manufacturer's instructions (Clontech).

10^5 cells Vero E6 were transduced with 5.10^6 TU NanoLg or NanoSm expressing lentiviruses for 4 hours, then cells were grown in DMEM 5% SVF + 1% Penicillin-Streptomycin + 8 µg/ml puromycin. Transduced cells were then amplified under puromycin treatment.

For the SARS-CoV-2 infection assay, Vero E6-NanoLg and Vero E6-NanoSm were plated separately at 25000 cells/well or co-cultured (12500 of each cell lines mixed/well) in white 96 well plates (Greiner Bio-One, #655083). One day after, cells were infected by serial dilutions of SARS-CoV-2 strain 1973 (hCoV-19/France/GES-1973/2020, EPI_ISL_414631) in 50 µl of DMEM, then incubated for 24 hours. Infection rates were measured by lysing cells after removal of the medium in 30 µl of 100-fold diluted NanoLuc substrate (Promega, #N1110). Luciferase enzymatic activity was measured using a CentroXS luminometer (Berthold; 2 s integration time).

Small Molecule Inhibition

GC376 (Sigma Aldric) and Spautin-1 (Calbiochem) were dissolved in DMSO to make a 100 mM stock solution. A 40 mM solution was prepared from the stock in DMSO, and then serial 2-fold dilutions were prepared in DMSO. The final range of inhibitors was from 40 mM to 0.63 mM (*i.e.* 8 2-fold dilutions), which is 200 times higher than the tested concentrations. Each compound dilution was diluted 1000-fold in DMEM (giving 2X concentrated chemicals with equal DMSO concentration for each)

Vero E6 NanoLg + Vero E6 NanoSm co-cultures (12500 each) plated the day before in 96 white plates were infected with 500 PFU of the previously described SARS-CoV-2 strain in 50 µl DMEM for 1 hour. Then 50 µl of the 2X concentrated compounds were added. Cells were incubated for 24 hours. Infection rate was measured by nanoluciferase activity reading as described above.

Each compound was measured in biological triplicates, each containing 3 to 8 technical replicates.

The results are presented as RLU ± standard deviations. Dose-response curves were plotted using Prism v9 (GraphPad) in which IC_{50} values were calculated with the variable slope model.
Cell viability was measured using the Cell-Titer-Glo® Luminescent Cell Viability Assay kit (Promega, G7750).

AlphaFold-Multimer *in silico* structure prediction

We produced predicted complex structures for the selected pairs of viral-host proteins by running the AlphaFold source code v2.3.2 (git commit hash 2e6a78f) from https://github.com/deepmind/alphafold in multimer mode with default parameters and with full databases via the provided Docker script. For each pair, we ran using input host protein sequence FASTA files from https://www.uniprot.org/ and generated 5 relaxed structures (from different random seeds) for each of the 5 multimer models. We kept the top-ranked relaxed structure in terms of model confidence in order to compute a predicted DockQ (pDockQ) score[25]. Following the CAPRI definitions for the correspondence between DockQ values and accuracy of predicted structures, we deem structures with pDockQ >= 0.23 as positive hits for the purposes of our *in silico* validation.

Figure legends

Fig 1. The physical core of the SARS-CoV-2-host interactome. (A) Physical core representing the SARS-CoV-2 interactions of our BioID dataset that have been identified by orthogonal, physical PPI detection methods. Bait proteins are represented as black nodes, the proximal host proteins by nodes color coded following the number of CRISPR hits as indicated in the color scale for the whole panel; **(B-E)** subnetworks of virus-host PPIs identified in our BioID dataset and involving host factors scoring in 3+ independent CRISPR screens, subdivided following their profile in the literature (**B**: In proximal and physical cores; **C**: in Physical core; **D**: in Proximal core; **E**: in Our BioID only).

Fig 2. SARS-CoV-2/human proximal network representation. (**A**) poly(I:C) remodeling of SARS-CoV-2 proteins proteins proximal partners; (**left**) graphical pipeline; (**middle**) distribution of proximal interactors changes for each viral bait protein. Horizontal axis depicts the percentage of the whole interactome of each bait significantly changed upon poly(I:C) treatment. Red and blue bars for increased and decreased proximal host partners, respectively. Number of unique host factors are shown on the graph. Bait names are colored according to their status, as depicted in the graphical pipeline; (**right**) Volcano plots representing one bait of each category (details are available in Supplemental table S6). (**B)** Global data-driven SARS-CoV-2/human proximal network projection on 2D and its relaxation upon removal of the host proteins targeted by eight or more viral baits. PxIs

stands for proximal interactions **(C)** Spatial localization of functional groups of proteins in the proximal network relative to the viral proteins (black nodes), as indicated. **(D)** Localization of human proteins essential for SARS-CoV-2 induced toxicity (3+ CRISPR hits) in the 3D SARS-CoV-2 /Host proximal interaction maps.

Fig 3. mN2H orthogonal validation of a subset of high confidence interactors. (A) Selection of the interactors and experimental pipeline for the mN2H validation. **(B)** Full SARS-CoV-2/Human contactome network identified in this study by mN2H.

Fig 4. mN2H interaction profiling. (A) The relative strength of PPIs, given by the distance of PPI to the positive threshold, is shown for host factors scoring positive with one or more viral factors. Their corresponding number of CRISPR Hits is indicated **(B)** Hierarchical clustering of the SARS-CoV-2 proteins and human host proteins basis of their mN2H interaction profiles. The heatmap represents the signal intensity derived from the relative strength of positive interactions (*i.e.* distance to the threshold). **(C)** Intra-virus interactions heat map, intraviral interactions were tested in triplicate using three different complementary NanoLuc configurations, the score corresponding to the numbers of positive interactions found. **(D)** Network representation of intraviral interactions. Thickness of the lines is relative to the distance from the threshold, the green proteins appear associated with organelles that are in close proximity to each other in the 3D network.

Fig 5. Organization and validation of the BioID and mN2H networks. (A) Snapshot of the 3D network depicting spatially resolved and colored clouds enriched in a subset of GO categories. Merged data from this study and the literature are graphically coded as described in the legend. **(B)** Violin plot of the viral-host distances in the 3D map according to the PPIs status, as described (Student t-test; *p*-values as indicated). **(C)** Modeling of the USP13/NSP13 complex and validation of USP13 as a host target**.** Molecular docking of USP13/NSP13 interaction. **(D)** Schematic of infection-induced cell fusion assay using nanoluciferase complementation. **(E)** Measured IC50 for Spautin-1 using the nanoluciferase complementation assay. Briefly, a co-culture of Vero E6-NanoLg and Vero E6-NanoSm, plated at equal density the day before infection, was infected at a MOI of 0.01 with the Wuhan strain of SARS-CoV-2. Increasing concentrations of Spautin-1 (red) and GC376 (blue) were added at the time of infection. Luciferase was measured 24 hours post-infection as a read-out of infection-induced syncytia formation.

Supplemental Figures legends

Fig.S1. Overview of the pipeline for identification of SARS-CoV-2/human interactome (A) Pipeline of the SARS-CoV-2-host interactome identification and potential mining.

Fig.S2. Expression, localization and biotinylation capacities of BirA*Flag-tagged SARS-CoV-2 proteins. Confocal imaging BirA*Flag proteins expressed in 293 Flp-In TREx in basal culture. Staining of DAPI, biotin (streptavidin- AlexaFluor488; proximal interactors) and anti-FLAG (anti-mouse AlexaFluor647; bait proteins).

Fig.S3. Expression of individual BirA* Flag-tagged SARS-CoV-2 proteins. N-ter and C-ter BirA* Flag-tagged SARS-CoV-2 expressing cells were lysed with Läemmli buffer, and western Blot were realized with an equivalent volume of whole cell extract. BirA* Flag-tagged proteins were revealed with anti-FLAG antibody (anti-mouse HRP secondary antibody), and anti-GAPDH (anti-mouse HRP secondary antibody) was used for protein loading control.

Fig.S4. Effect of individual BirA* Flag-tagged SARS-CoV-2 proteins on IFN activation. HEK293-ISRELuc reporter cells were co-transfected with BirA*Flag alone or the indicated BirA*Flag SARS-CoV-2 proteins and **(A, B)** MAVS expression plasmid **(C, D)** pcDNA3. IFN was measured as luciferase (RLU) Fold change over ISRELuc activity without MAVS co-expression.

Fig.S5. Metascape analysis of the enriched targeted categories per viral protein. Metascape Express Analysis report was used to automatically select the functional categories enriched for each viral. The detailed output of the functional enrichment analysis, including the gene list assigned to each category for each bait protein, is provided in Supp Table 4.

Fig.S6. Metrics and features of the BioID-based network and 3D map. Analysis of the effect of filtering out host proteins with BioID degree greater than seven on the network topology: **(A)** Histogram of host protein degree and dependence of the network's total number of nodes and edges. The majority of host proteins have one or two viral interactions. After applying the degree cutoff, around 82% and 53% of nodes and edges respectively are retained. **(B)** Histogram of viral protein degree before and after filtering for highly specific interactors. We observe that the specificity ratio is generally higher for viral proteins with smaller total number of interactions; Distance-based analysis of the network's 3D representation: **(C)** Host target density and total population dependence with radial distance from dense core center for the full and filtered (host protein degree <= 7) networks. Identification of dense core size via the radial location of the density interpolated line inflection point for the case of the filtered network. Host proteins below the cutoff are considered to belong inside the dense core. **(D)** Distance histograms, densities and averages (vertical dashed lines) for all viral-host protein pairs tested in mN2H, including breakdown by mN2H +/- status, and subset of BioID+ interactions. mN2H+ positive pairs are found at lower average distances, their distribution deviating significantly from the overall mN2H matrix distribution and showing similarity to the BioID+ distribution. **(E)** Distance histograms, densities and averages

(vertical dashed lines) restricted to mN2H-tested BioID+ protein pairs including breakdown by mN2H+/- status. Despite having no influence on the construction of the 3D layout, mN2H+ pairs exhibit a lower average distance, reinforcing the predictive power of the map; **(F)** Distance histograms, densities and averages (dashed lines) restricted to mN2H-tested BioID- protein pairs and breakdown by mN2H+/- status. Despite having no influence on the construction of the 3D layout, N2H+ pairs exhibit a lower average distance, reinforcing the predictive power of the map.

Fig.S7. AlphaFold-Multimer predicted structures of SARS-CoV-2-host complexes. Cartoon representation of the SARS-CoV-2 /Human protein complexes with a pDockQ score >0.23. The viral proteins are shown in light violet, the host proteins in light pink. Protein surfaces are shown. When no structure prediction was available, the resulting unstructured representations obtained by AlphaFold were not shown for the corresponding proteins, owing to the lack of information they bring. The *.pdb files corresponding to each predicted complex are provided as hypertext links in **Supp table 11**.

Fig.S8. Benchmarking of the Nanoluciferase auto-complementation system for SARS-CoV-2 infection measurement. Vero E6-NanoLg and Vero E6-NanoSm, either cultured separately or co-cultured, equal density were infected with serial dilutions of the Wuhan strain of SARS-CoV-2. Luciferase was measured 24 hours post-infection as a read-out of infection-induced syncytia formation.

Supplemental tables

Supp. Table 1. List and sequences of SARS-CoV-2_{wuhan} ORFs used in this study. Each gene synthesize cDNA encoding viral proteins from the early SARS-CoV-2 strain WuhanHu-1 (GenBank MN908947 [50]) where obtained from[34] as gateway pENTRY clones and transferred in the BioID or mN2H-dedicated plasmides

Supp. Table 2. SARS-CoV-2 BioID High data. Tab A. Full high-confidence proximal interactors. The number of baits detecting a given prey protein is indicated in the 'Indegree' column. The 'Abundance vs background' column represents the ratio between the detected prey in given bait assays against the control group (see Tab for RawQuant Perseus data). The 'poly(I:C) vs. basal' shows the log2 fold LFQ intensity changes between the preys detected after poly(I:C) treatment enrichment and the corresponding basal condition. Over 1 is considered as either enrichment or gain (if absent from the basal condition), between 1 and −1 corresponds to unchanged, and below −1 shows decreased or lost hits. When no value is indicated, it corresponds to interactions not detected in the poly(I:C) and the corresponding basal condition. The reported interactions columns display the bait-prey pairs previously identified in the literature in other proximal interactomics or orthogonal physical PPIs detection studies (see Tab C-1 and C-2 respectively for details and references). For each cellular protein, the

number of CRISPR screens where they scored as necessary for SARS-CoV-2 infection is indicated. Tab B. Raw output of the Perseus analysis of MaxQuant LFQ results. Tab D ToppCluster analysis for enriched/ gained interactors. Tab E. ToppCluster analysis for enriched/ gained interactors after poly(I:C) treatment. Tab F. Reported interactions between the viral proximal factors

Supp. Table 3. Overlap of our SARS-CoV-2 BioID Datasets with reported SARS-CoV2 interactomes

Supp. Table 4. Functional annotation of high confidence proximal interactors for each bait viral protein . Functional annotations are generated using a compiled list of proximal interactors from N-ter, C-ter and poly(I:C) conditions.

Supp. Table 5. GO-term analysis of the BioID dataset compared with dataset combining all other interactomes

Supp. Table 6. Poly:IC remodeling of SARS-CoV-2 proximal interactomes. Fold change of proximal interactions is given for each host factor per viral protein, and represented by volcano plots.

Supp.Table 7. Factors of the dense core (less than 0.3 of the centroid in the 3D map coordinates)

Supp.Table 8. mN2H raw data and normalization. Mean RLU of quadriplicates (Fold Change or FC matrices) or duplicates to quadriplicates (CRISPR matrices) of mN2H PPIs are given along with the normalization pipeline for mN2H positive signals scoring

Supp. Table 9. Summary of the mN2H data. The distance to the positive PPIs threshold is given for each virus-host PPIs tested. PPIs detection in previous proximal interactomics and physical interactomics studies is given, as well as the number of CRISPR Hits of the host factor.

Supp. Table 10. Intra-virus mN2H raw data and normalization.

Supp. Table 11. Predictions of SARS-CoV-2-host complexes. pDockQ scores (Tab A) and confidence scores (Tab B) of mN2H positive virus-host PPIs obtained by AlphaFold-Multimer. pdb files of the corresponding predicted virus-host complex structures are accessible through hypertext links.

Supp. Table 12. Raw data of the IFN activation signaling pathway upon expression of SARS-CoV-2 proteins.

Acknowledgments

The operational costs of this project were supported by the *Agence Nationale pour la Recherche* (ANR) Flash COVID-19 funding scheme (DARWIN project), the Task Force Covid19 Institut Pasteur, RECOVER project funded by the European Union's Horizon 2020 (grant agreement No. 101003589) , Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant n°ANR-10-LABX-62-IBEID), the *Pasteur-Weizmann Council*, the French Government's Investissement d'Avenir programme and I-Site Lille.

Métropole Européenne de Lille (France) Accueil de Talent (MONET) supported E.MN.L.

G.D. is supported by a doctoral fellowship from Université Paris Cité.

A.K. and Y.R. were supported by Task force COVID-19 funding (SARS-CoV-2 sensing) and *Agence Nationale pour la Recherche* (ANR) Flash COVID-19 funding scheme (SARS-Cov-2 immunRNAs project).

Y.S. was supported by the Operational Program Competitiveness, Entrepreneurship and Innovation, NSRF 2014-2020, Action code: MIS 5002562, co-financed by Greece and the European Union (European Regional Development Fund) and by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 838018.

G.A.P. was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I) under the "First Call for H.F.R.I Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (grant 1855-BOLOGNA).

A.-C.G. holds the Canada Research Chair (Tier 1) in Functional Proteomics.

P.A. and P.F.B. were funded by the European Research Council's Horizon 2020 Research and Innovation Action (Grant Agreement 101003633 - RiPCoN).

E.C. was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 843052.

We would like to thank Antoine Touzé, Yves Gruel and Mathias Faure for sharing their expertise in SARS-CoV-2 related processes. Dushyandi Rajendran kindly provided technical assistance to obtain the BioID vectors. We acknowledge Marc Vidal, David and Michael Calderwood for their fruitful comments and inputs in the framework of VirHostOme consortium. We thank Maxence Wisztorski for technical assistance with MS data, Irène Gadotti and Nathalie Vasseur for their logistic support.

Author contributions

E.MN.L., P.S-T., A-C.G., E.C., Y.J. and C.D. obtained and cloned the SARS-CoV-2 polypeptides coding sequences. E.MN.L. and P.S.T. produced BioID vectors and E.MN.L. conducted all BioID sample purification. E.C. realized the fluorescence microscopy experiments and cell cultures for the BioID part. J-P.G. ensured all mass spectrometry acquisition and contributed to the data analyses conducted by E.C. K.B., Y.R. and A.K. realized the quality controls and luciferase activity assays. G.D., M.A., M.D.S. realized the mN2H experiments. F.D., F.A., S.M., C.D. set up the novel luciferase-based infection assay, F.D. and C.D. performed the follow-up experiments. Y.S., N.S. and P.A. performed the structural modeling analyses. N.P. and S.V. ensured the bioinformatic statistical analyses. P.F-B., P.A., A-C.G., S.VD.W., Y.J., Y.S., C.D. and E.C. contributed to the project development through the VirHostOme consortium assembled *ad hoc* by P.F.-B to aggregate international interactomics research efforts. E.C., Y.J., C.D., Y.S. and E.MN.L conceived the project. C.D. and E.C. secured funding. Y.S. conceived and coded the online 3D network interactive visualization interface, with inputs from E.C., C.D. and Y.J. and support of G.A.P. E.C., C.D., Y.S., M.A., A.K. and G.D assembled the figures and wrote the paper, with scientific and editorial inputs from Y.J., P.F-B. and E.MN.L.

Declaration of interests

The authors declare no competing interests.

Fig.1

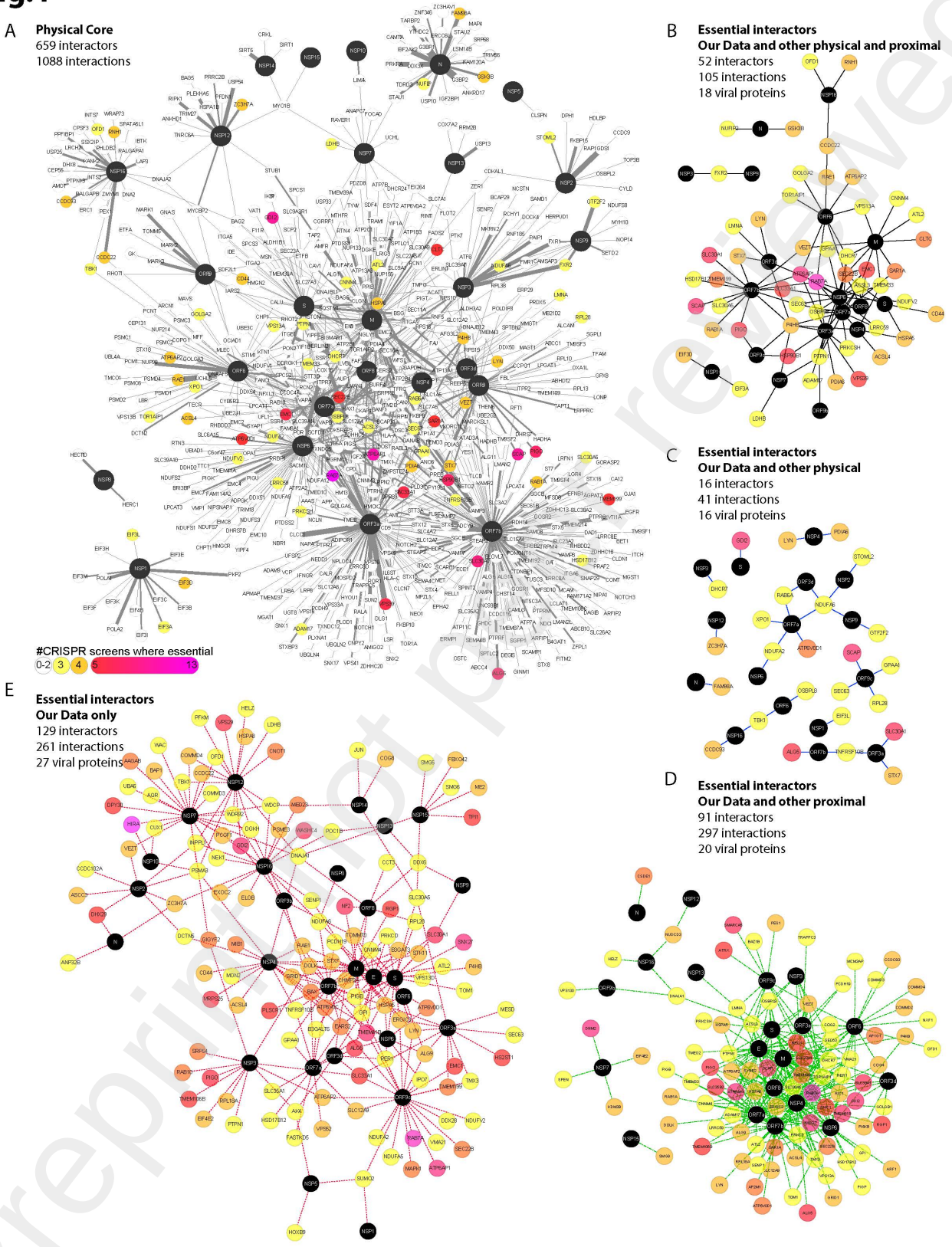
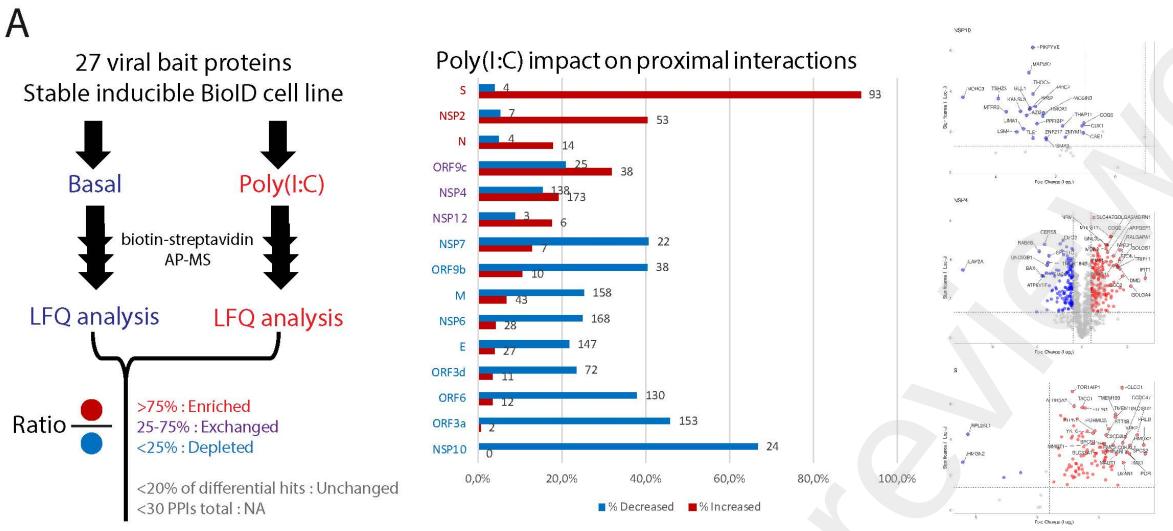
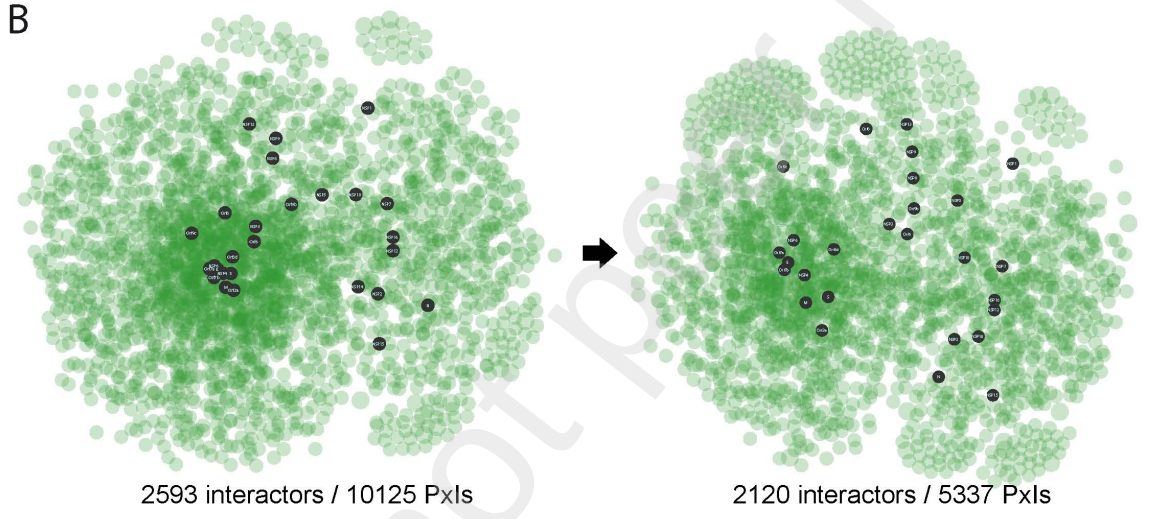


Fig.2





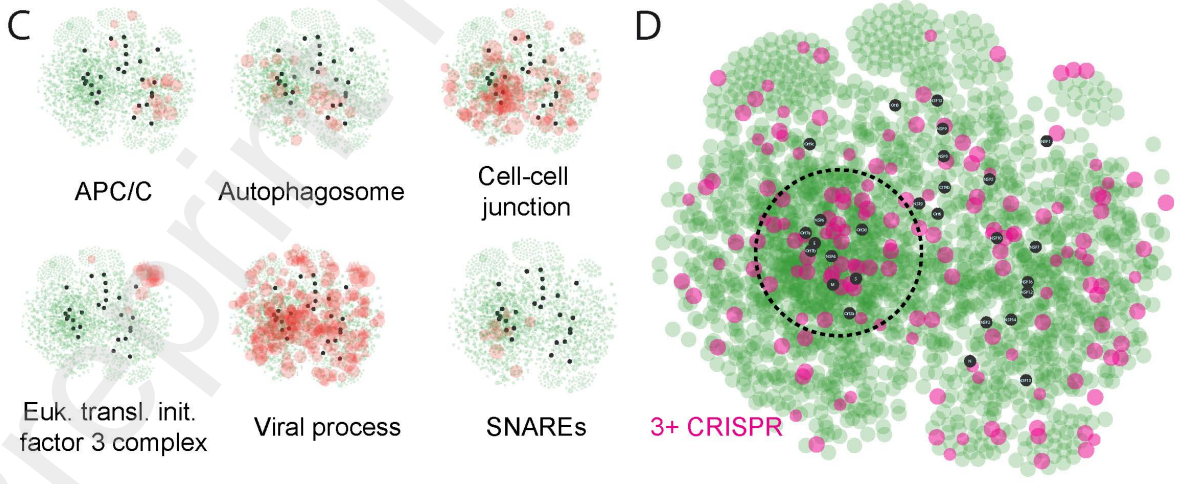
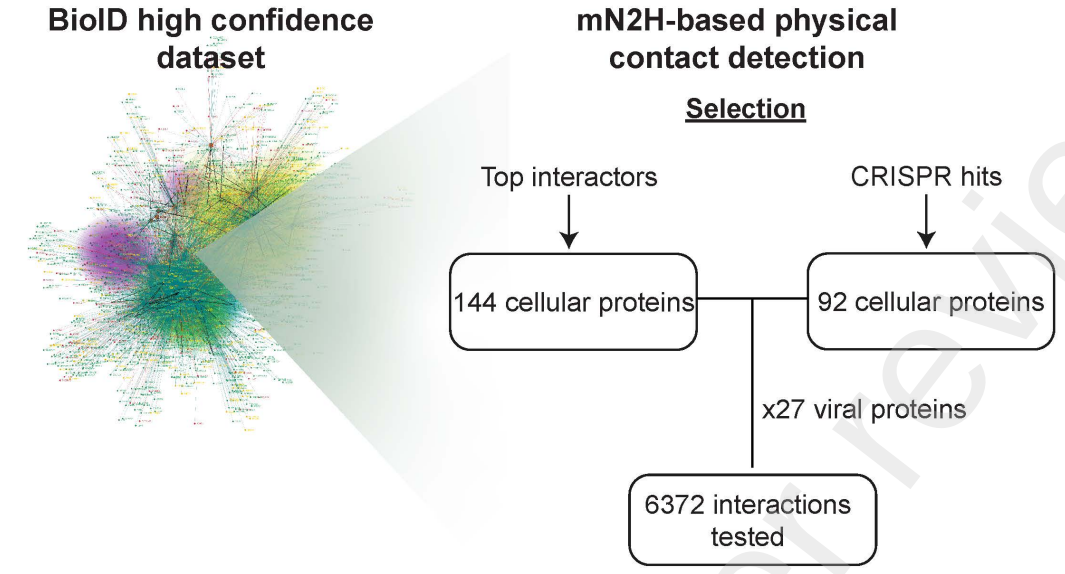
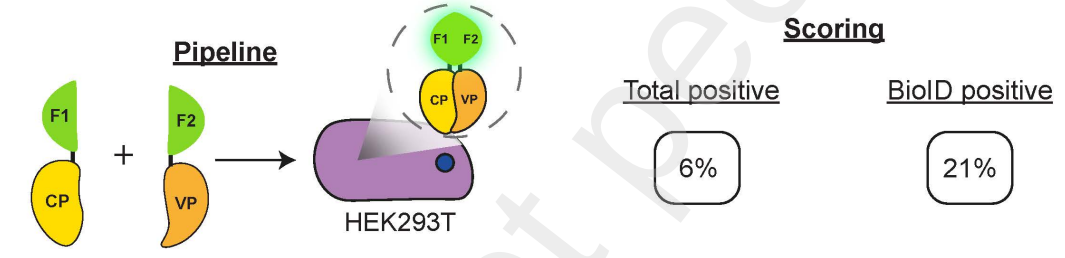


Fig.3

A





B

mN2H-positive interactions
414 PPIs
25 viral proteins
118 host factors

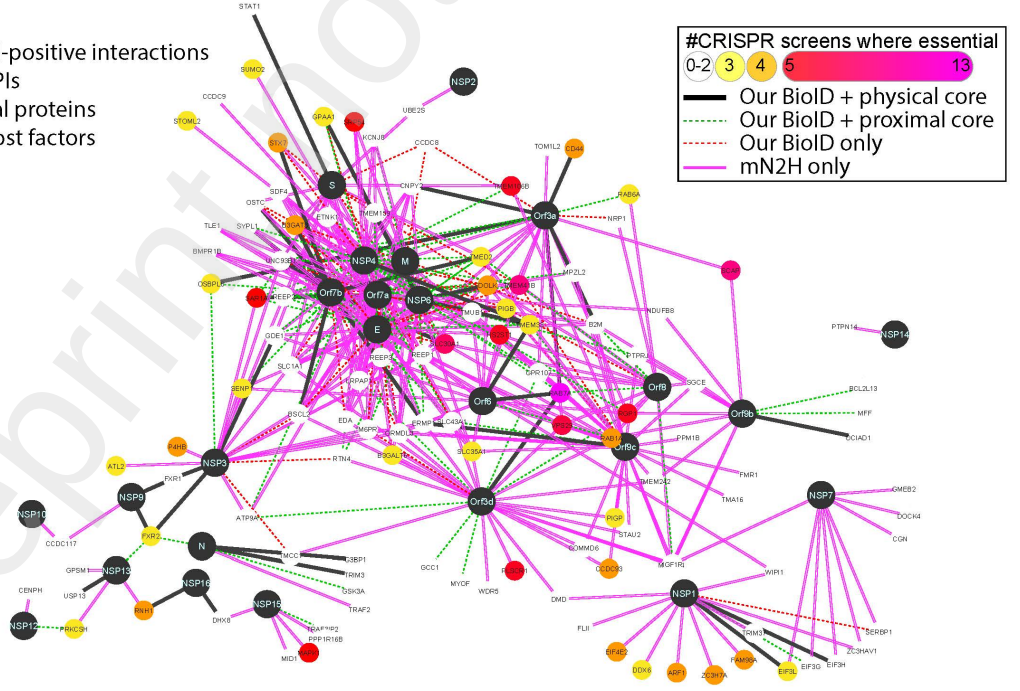


Fig.4

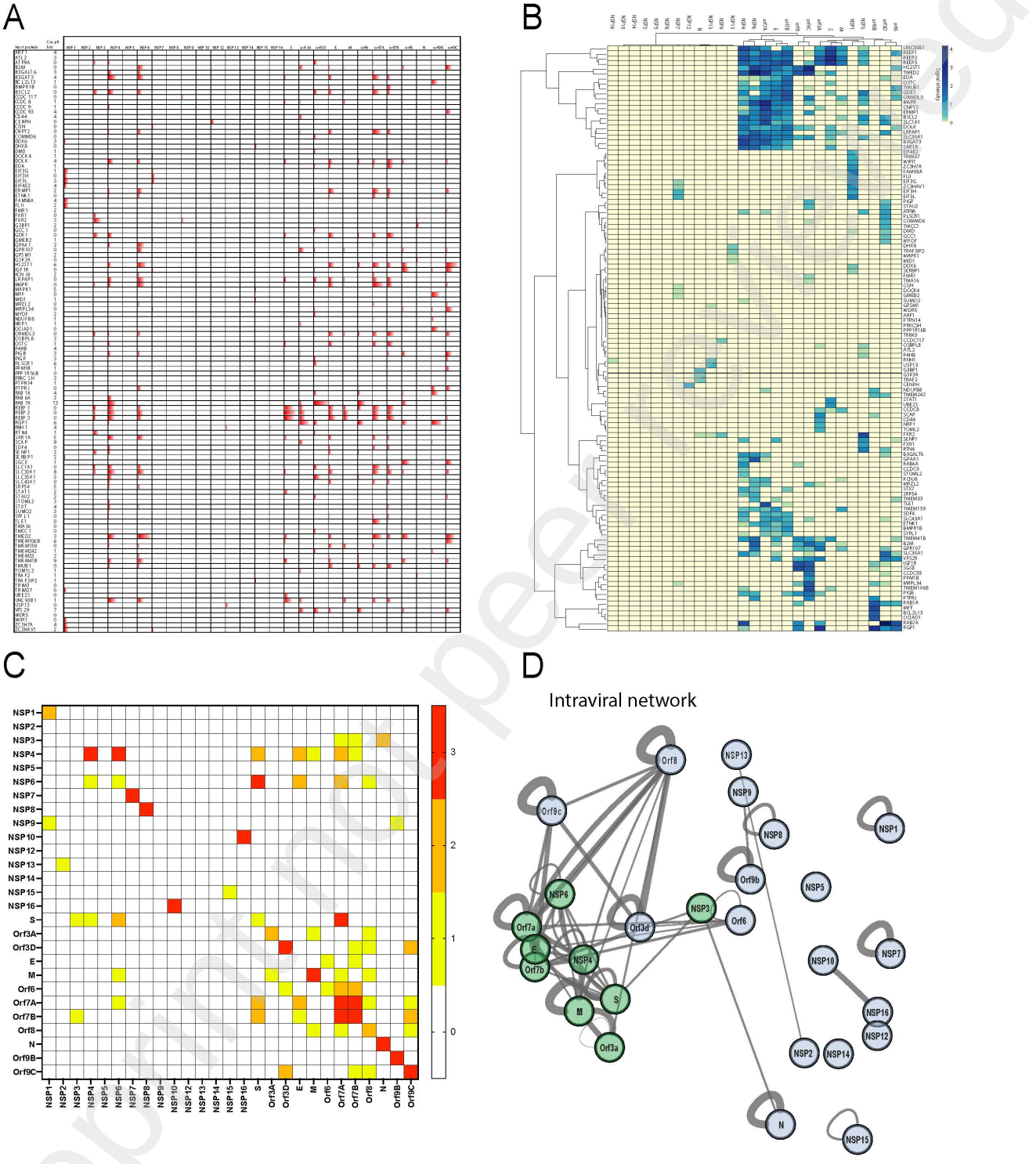


Fig.5