DATABASE
The Journal of Biological Databases and Curation

Database, 2020, 1–10
doi:10.1093/database/baaa096
Database tool

Database tool

The IMEx coronavirus interactome: an evolving map of *Coronaviridae*–host molecular interactions

L. Perfetto[1,†], C. Pastrello[2,†], N. del-Toro[1], M. Duesbury[1,3], M. Iannuccelli[4], M. Kotlyar[2], L. Licata[4], B. Meldal[1], K. Panneerselvam[1], S. Panni[5], N. Rahimzadeh[3,6], S. Ricard-Blum[7], L. Salwinski[3], A. Shrivastava[1], G. Cesareni[4], M. Pellegrini[8], S. Orchard[1], I. Jurisica[2,9], H. Hermjakob[1] and P. Porras[1,*]

[1]European Molecular Biology Laboratory, Wellcome Genome Campus, European Bioinformatics Institute (EMBL-EBI), Hinxton, CB10 1SD, UK, [2]Krembil Research Institute, Data Science Discovery Centre for Chronic Diseases, University Health Network, 5KD-407, 60 Leonard Avenue, Toronto, ON, M5T 0S8, Canada, [3]UCLA-DOE Institute, UCLA, Los Angeles, CA 90095, USA, [4]Department of Biology, University of Rome Tor Vergata, Via della Ricerca Scientifica, Rome, 00133, Italy, [5]Department of Biology, Ecology and Earth Sciences, Università della Calabria, Rende, 87036, Italy, [6]Providence John Wayne Cancer Institute, Department of Translational Molecular, Santa Monica, CA 90404, USA, [7]Univ Lyon, University Claude Bernard Lyon 1, INSA Lyon, CPE, Institute of Molecular and Supramolecular Chemistry and Biochemistry (ICBMS), UMR 5246, F-69622 Villeurbanne, 69622, France, [8]Department of Molecular, Cell and Developmental Biology, UCLA, Los Angeles, CA 90095, USA and [9]Departments of Medical Biophysics and Computer Science, University of Toronto, Toronto, ON, M5T 0S8, Canada

[*]Corresponding author: Telephone +44 1223 494482; Email: pporras@ebi.ac.uk

[†]These authors contributed equally to this work.

Citation details: Batra, D., Diwan, N., Upadhyay, U. *et al.* The IMEx coronavirus interactome: an evolving map of *Coronaviridae*–host molecular interactions. *Database* (2020) Vol. XXXX: article ID baaa096; doi:10.1093/database/baaa096

Received 16 June 2020; Revised 19 October 2020; Accepted 23 October 2020

Abstract

The current coronavirus disease of 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus (SARS-CoV)-2, has spurred a wave of research of nearly unprecedented scale. Among the different strategies that are being used to understand the disease and develop effective treatments, the study of physical molecular interactions can provide fine-grained resolution of the mechanisms behind the virus biology and the human organism response. We present a curated dataset of physical molecular interactions focused on proteins from SARS-CoV-2, SARS-CoV-1 and other members of the *Coronaviridae* family that has been manually extracted by International Molecular Exchange (IMEx) Consortium curators. Currently, the dataset comprises over 4400 binarized interactions extracted from 151 publications. The dataset can be accessed in the standard formats recommended by the Proteomics Standards Initiative (HUPO-PSI) at the IntAct database website (https://www.ebi.ac.uk/intact) and will be continuously updated as research on COVID-19 progresses.

Introduction

Severe acute respiratory syndrome, or SARS, emerged as a life-threatening viral disease of unknown origin in late 2002 in the Guangdong Province of southern China caused by the severe acute respiratory syndrome coronavirus (SARS-CoV) (1). The SARS-CoV-2 is a related virus responsible for the current outbreak of coronavirus disease of 2019 (COVID-19) (2). As of September 2020, over 26.5 million people globally have been shown to be infected with the virus, with more than 873 000 deaths directly attributed to its effects (worldometer, https://www.worldometers.info/coronavirus/). To distinguish the strains that caused the SARS outbreak of 2002 from other SARS-related, nonhuman-hosted SARS-CoV strains, we refer to the former as SARS-CoV-1.

The COVID-19 pandemic has resulted in massive scientific efforts attempting to fight the disease and understand the biology of the virus. This has resulted in enormous challenges for the research scientist when attempting to find and select information relevant to specific areas of viral biology and pathology. In order to aid the scientific community and expedite drug and vaccine development, multiple data curation efforts have been undertaken to perform a critical assessment of the literature and represent different aspects of the virus and the disease in a structured and computationally accessible manner. One recent example of such efforts is the COVID-19 Disease Map (3), a community effort to capture the intricate aspects of SARS-CoV-2 biology as reusable and interoperable pathway maps, so they can be used in systems biology and modeling pipelines.

Identification of virus–host interactions and the analysis of the topological structure of a relevant molecular interaction network are the necessary steps to understand the cellular mechanisms involved in a biological process such as viral infection of a cell. A detailed map of the interactions between human and pathogen proteins will aid in a more complete awareness of the mechanisms of infection and subsequent viral replication, assembly and release and may help to identify novel drug targets, or assist in a rapid and more accurate repurposing of existing drugs for treating or preventing infection. Further to this, such networks can be used to study changes in the transcriptome or proteome of a virally infected cell when compared to normal cells. Co-regulated genes or proteins that also co-cluster in this network may indicate that these entities are involved in the same biological process or are members of the same functional complex.

For such a network to be of value to the researcher, a certain amount of metadata needs to be provided, which enables the assessment of network quality and data types. These data need to be supplied in a standardized and computer-accessible format, which allows for scoring, filtering and selection. The International Molecular Exchange (IMEx) Consortium (4) has been providing such data for over 15 years, supplying experimental details using controlled vocabulary terms, captured using a detailed curation model. Important aspects of an interaction experiment are described, including host organism, interaction detection and participant identification methodologies as well as details of the constructs such as binding domains and the effects of site-directed mutations. Current membership of the IMEx Consortium includes the IntAct (5), MINT (6), DIP (7), UniProt (8), MatrixDB (9) and IID (10) data resources, who collaborate to provide the users with a single, consistent viral–host dataset to work with. When a novel virus emerges, as was the case with SARS-CoV-2 in 2019, the study of closely related species, such as SARS-CoV-1 and other coronaviruses, may help with this process, giving the scientific community time to produce species-relevant data. For this reason, the network includes data on all coronaviruses available in the scientific literature. While primarily consisting of protein–protein interaction data, the network also contains interactions with lipids, glycosaminoglycans and RNAs, again curated to IMEx standards. The data are open under a CC-BY 4.0 license and downloadable from the IntAct website and FTP in PSI-MI standard tab-delimited and XML-based formats. A brief description can be found at http://www.ebi.ac.uk/intact/resources/datasets, and a collection of interactive network representations of the dataset at the time of writing is available at http://www.ndexbio.org/#/networkset/4c2268a1-a0f0-11ea-aaef-0ac135e8bacf. The dataset will be expanded and updated with every IntAct release.

Methods

Data sources

All statistics and details shown on this manuscript are derived from the IntAct database release of 2020-07-11. Complex Portal annotations used in Supplementary Table 4 were also obtained from the same data release.

Specific analyses include the following datasets: SARS-CoV-2—all human targets of SARS-CoV-2 viral proteins; SARS-CoV-1—all human targets of SARS-CoV-1 viral proteins; Gordon_LT—all human targets of SARS-CoV-2 viral proteins derived by Gordon *et al.* plus selected SARS-CoV-2 low-throughput (LT) studies; Li_LT—all human targets of SARS-CoV-2 viral proteins derived by Li *et al.* plus selected SARS-CoV-2 LT studies and Stukalov_LT—all human targets of SARS-CoV-2 viral proteins derived by Stukalov *et al.* plus selected SARS-CoV-2 LT studies. The selected LT studies added to Gordon, Li and Stukalov studies represent interactions that were detected in studies with less than 50 interactions and not detected in the high-throughput (HT) ones. Such interactions involve SARS-CoV-2 Spike (interacting with seven human proteins, such as ACE2, BSG, DEFA5, FURIN, CCNB1, CLEC4M and CD209), nsp3 (interacting with human PARP10) and nsp8 (interacting with human HLA-A). Formatted subsets are available in Supplementary Table 5. We also included in Supplementary Table 1D a reduced version of the dataset where only *Coronaviridae* protein interactions are included and interactions with human proteins are highlighted.

Analysis software and packages

Analyses were done using R 4.0.0 (11) and the R data.table package (12). Venn diagrams were created using ggVennDiagram 0.3 (13). All other plots were created using ggplot2 3.3.0 (13) and wesanderson (14) packages. The network in Figure 1 was created using NAViGaTOR 3.0.13 (15); the data used to create the network are provided in Supplementary Table 5F. The rest of the interactive networks were created using Cytoscape 3.8.0 (16), and all networks were uploaded to NDEx using Cytoscape.

Tissue enrichment analysis

Analysis was performed using TissueEnrich (17) 1.8.0 R package on human protein targets using Protein Atlas (18) expression data. Background was the entire proteome as listed in the UniProtKB website (https://www.uniprot.org/proteomes/UP000005640). Only hits with a fold change above zero are shown. Log10 *P*-value was zero for all tissues. Full data are available in Supplementary Table 3.

Pathway enrichment analysis

Pathway enrichment analysis was performed using pathDIP 4 (19) API in R, with literature curated set. Only pathways with false discovery rate (FDR) < 0.01 (Benjamini Hochberg (BH)-method) were considered. As the majority of enriched pathways (75% for SARS-CoV-1, 80% for SARS-CoV-2, 72% for Gordon_LT, 76% for Li_LT and 55% for Stukalov_LT) were from Reactome database (20), we further organized them using level-one Reactome ontology to create the figure. All enriched pathways are listed in Supplementary Table 2, and pathways present in multiple sets are highlighted in the 'overlap' tab. For network annotation, each protein was linked to the pathway it is part of with the lowest FDR (BH-method) and annotated with level-one Reactome ontology according to such pathway. When a protein was annotated with two different level-one terms (this happened only for 58 targets in common between SARS-CoV-1 and SARS-CoV-2), only the SARS-CoV-2 level-one term was used.

Results and discussion

As of August 2020, the dataset contains 3735 unique interacting molecule pairs, represented in 4479 binarized interactions extracted from 151 publications, 8 of which are preprints from bioRxiv. The entire 'coronavirus dataset' can be downloaded from the FTP site in PSI-MI standard XML-based formats: PSI-MI XML 2.5 and 3.0 (format documentation available at https://psicquic.github.io/PSIMIXML.html). Each file corresponds to one publication in the coronavirus dataset. In addition to the XML format, the data can also be browsed on the IntAct webpage (www.ebi.ac.uk/intact/query/annot:"dataset:coronavirus"), where it is available for download in additional formats, such as the tab-delimited PSI-MI-TAB 2.7 (format documentation available at https://psicquic.github.io/MITAB27Format.html). Viral specific interactomes can also be retrieved using MIQL syntax. For example, to filter for human-SARS-CoV-2 interactions, the string 'taxid:2697049 AND taxid:9606' can be used in the search box.

NDEx archive

A snapshot of the coronavirus dataset hosted in IntAct at the time of writing is also available as a collection of interactive network representations in the NDEx archive. Here, we submitted five networks, each highlighting a key aspect of the dataset:

- Full dataset in collapsed view, with unique molecule pairs linked by edges representing aggregated evidence: https://doi.org/10.18119/N9MP4S
- Full dataset and full evidence details represented as multiple edges: https://doi.org/10.18119/N9RC8F
- Full dataset, highlighting evidence where mutant forms of proteins involved are available: https://doi.org/10.18119/N9W590
- Full dataset, highlighting evidence where experi mental information about binding regions of proteins involved are available: http://doi.org/10.18119/N90W3W
- Subset of SARS-CoV-1 and SARS-CoV-2 to human targets interactions with pathway annotation (Figure 1): https://doi.org/10.18119/N9831P.

Details about the visual styles used to depict evidence, species and further information are provided in the NDEx

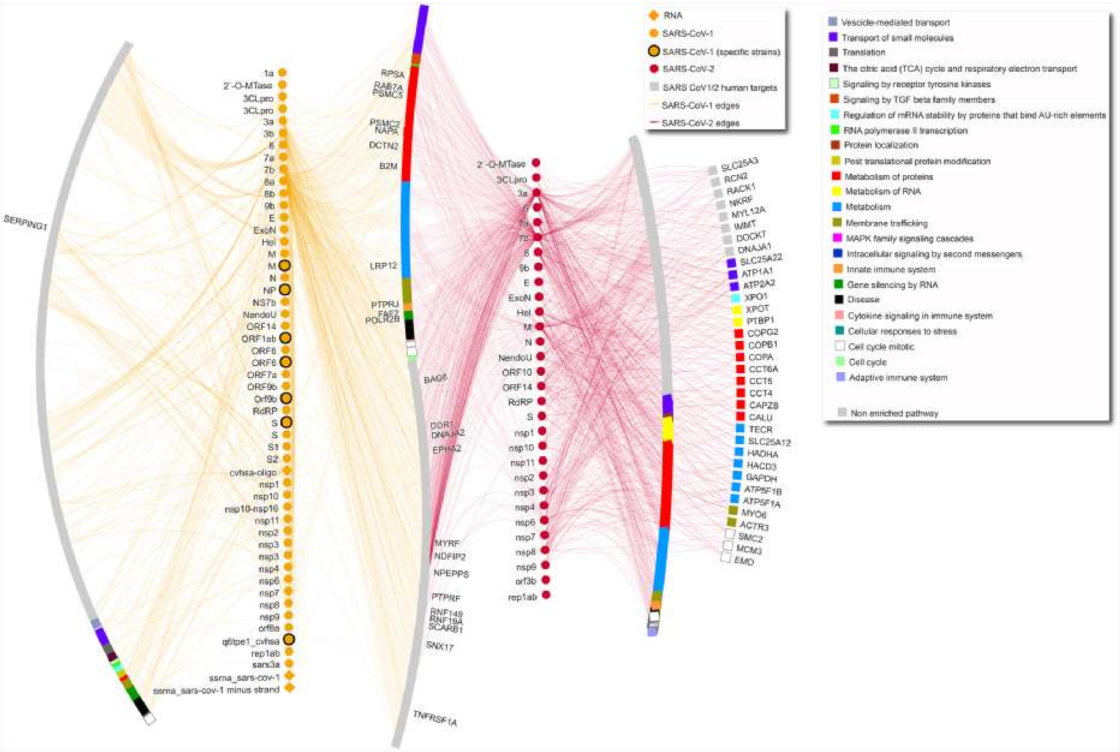

Figure 1. SARS-CoV-1–human and SARS-CoV-2–human network. The network includes only interactions between SARS-CoV-1/2 proteins and human targets. Human targets are annotated with level-one Reactome ontology, as shown in legend. Only targets with degree equal or higher than 5 are labeled. Diamond SARS-CoV-1 nodes represent RNA molecules. SARS-CoV-1 specific strains include strains PC4-145, Frankfurt 1, Urbani, TJF, HB and GZ02. Gray nodes are annotated with pathways that are not enriched in this network. In the network, the leftmost arc shows human proteins targeted only by SARS-CoV-1 proteins, the vertical line made of yellow nodes indicates SARS-CoV-1 proteins, the S-shaped arc shows human proteins targeted both by SARS-CoV-1 and SARS-CoV-2 proteins, the vertical line made of red nodes indicates SARS-CoV-2 proteins, and the two right arcs show proteins targeted only by SARS-CoV-2, with the rightmost arc including only proteins with degree equal or higher than 5.

network description and in Supplementary 'methods' section.

Other access options

Using the same search strings as on the IntAct website, the data can also be accessed via the PSICQUIC (21) API, the IMEx Consortium webpage (www.imexconsortium.org) and the VirusMentha (22) browser at https://virusmentha.uniroma2.it/.

As with all data hosted by IntAct, the dataset is available under the CC BY 4.0 license

The complete dataset includes all the interactions that have been found in publications that fulfill the criteria reported in the next paragraph. The data refers mostly to protein–protein interactions (3568 interactions), plus some interactions involving different types of RNA (84 interactions) or small molecules (57 interactions). While data on 81 organisms are included (Supplementary Table 1A), most interactions refer to SARS-CoV-2–human and SARS-CoV-1–human interactions (2019 and 870 unique interactions, respectively). The dataset of only human–*Coronaviridae* interactions is highlighted in Supplementary Table 1D. Figure 1 shows SARS-CoV-2– and SARS-CoV-1–human interactions, separating human targets that are specific for one or the other coronavirus or in common between the two, also highlighting affected pathways.

IMEx Consortium curators have collated interaction evidence from scientific articles and preprints using the following selection criteria:

(i) The publication contains interactions involving proteins from any virus member of the *Coronaviridae* family (NCBI taxon ID 11118). This includes not only SARS-CoV-1 and SARS-CoV-2, but also Middle East Respiratory Syndrome (MERS)-CoV and members of the family that infect other mammals.

(ii) The publication contains interactions of human proteins with established relevance for SARS-CoV-2 life cycle (e.g. ACE2 interactions have been included in the dataset).

(iii) Every interaction described in these publications is curated and included in the dataset, even if their relevance to COVID-19 might seem limited. This results in the inclusion of data from apparently irrelevant species (e.g. yeast) that is of interest from the phylogenetic and evolutionary point of view.

(iv) Preprints are considered and, if deemed appropriate, curated when containing SARS-CoV-2 data, an exception to IMEx practice of representing only peer-reviewed research. This reflects the interest these data have engendered during the pandemic. These datasets are clearly marked as prepublication and will be re-curated, if necessary, when published.

The IMEx curation model captures the details of specific constructs used for the detection of interactions. This allows for the representation of different construct-associated features, such as specific mutations affecting interaction outcome (23) or sequence regions that are associated with binding. Most of these in-depth studies are centered around the Spike–ACE2 interaction for SARS-CoV-2 and SARS-CoV-1. The only other SARS-CoV-2 mutation reported so far is Nsp5/3C-like proteinase p.Cys145Ala catalytically inactive mutant, which exhibits an interaction profile different from the canonical form in an affinity purification and mass spectrometry (AP-MS) study (24). There are a few other mutations reported for SARS-CoV-1 and in other members of the *Coronaviridae* family, but it is clear that specific variation effects on coronavirus-related interactions is an area requiring extensive exploration. As shown in Figure 2A, fragment constructs have been used in more studies than mutations, although some of these might have been designed for convenience reasons (e.g. constructs that are easier to express in heterologous systems). Interactive network representations highlighting mutations and binding regions can be found in https://doi.org/10.18119/N9W590 and in https://doi.org/10.18119/N90W3W.

Regarding interaction data generation approaches, the experimental setup used to detect an interaction is summarized by three key fields in the IMEx curation model: the method used to determine that an interaction is happening ('interaction detection method'), the method used to determine which molecules are involved ('participant detection method') and the biological environment in which the interaction occurs ('host organism'). Each of these is described by an appropriate controlled vocabulary term, making the data readily searchable. In the coronavirus dataset, the overwhelming majority of data were generated by AP-MS approaches performed in A549, HEK293(T) or HeLa cells (Supplementary Table 1B). This type of data produces sets of potential interacting partners (preys) associated with a bait of interest but does not directly identify binary interacting partners. Also, the data needs to be automatically expanded into binaries for its representation in tabular or network graph form. This accounts for a large proportion of binary interactions from spoke-expanded n-ary relationships found in the dataset: 3267 out of 4477 interactions (73%) are expanded n-ary relationships, while only 41% of the binary interactions found in the full IntAct database are expanded n-ary relationships.

Most content has been curated after declaration of the COVID-19 pandemic on 11 March 2020 (Figure 2). The data growth timeline shows four jumps in interaction numbers, three for SARS-CoV-2 and one for SARS-CoV-1, due to curation of HT studies not available for other *Coronaviridae* (Figure 3A). This is also reflected in the pattern of data growth regarding detailed information about mutations and binding regions (Figure 2A). The plots also illustrate that at this point there are more studies dealing with detailed interaction data for SARS-CoV-1, a situation that will likely change as more SARS-CoV-2 studies are performed.

Interaction data are derived from both, small-scale studies focused on one or just a few interactions (LT) and large-scale screenings able to detect hundreds or thousands of interactions in a single experiment (HT). The extreme urgency in the study of SARS-CoV-2 has resulted in a strong dominance of HT data for this species in comparison with the other members of *Coronaviridae* reported in the dataset (Figure 3A). Additionally, SARS-CoV-2 small-scale studies are mainly focused around the ACE2–Spike interaction due to its relevance for virion recognition and infection. For the remaining interactions, most SARS-CoV-2 data comes from three studies: Stukalov *et al.* (25), where A549 cell lines were transduced with viral proteins followed by AP-MS analysis, and Gordon *et al.* (24) and Li *et al.* (26), both focused on AP-MS detection of interacting candidates in HEK293(T) cells transfected with SARS-CoV-2 proteins. The studies are distinct, showing small overlap (Figure 3B–D) with each other and no overlap at all with small-scale studies. They also show different degree distribution patterns (Figure 3E). Gordon *et al.* shows an unexpected pattern of fully isolated components, likely due to stringent target selection.

Lack of overlap between different HT interaction datasets is a long-recognized phenomenon, since different experimental approaches are better suited to detect interactions featuring specific physicochemical characteristics and protein abundances, among other parameters (27–29). Even in this case, where the approach used is broadly similar, they likely reflect methodological differences. AP-MS datasets are very sensitive to protein abundances and affinities, as well as to the selection and orientation of tags and

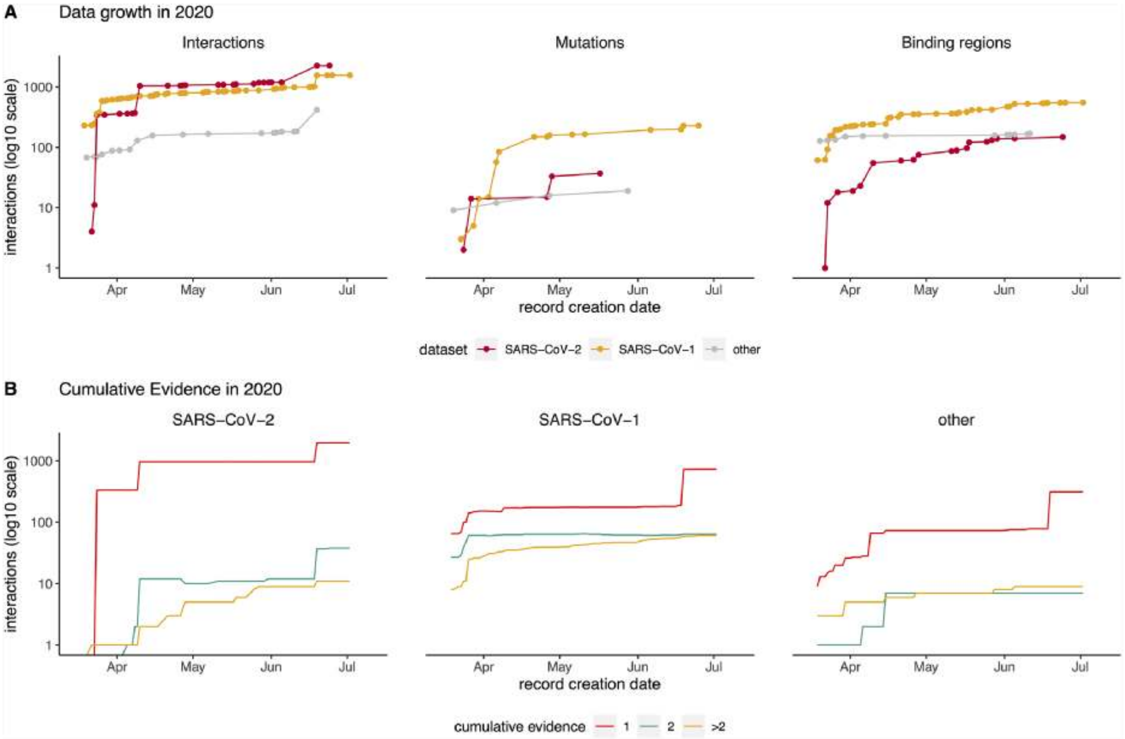

Figure 2. Timeline showing data captured in IMEx resources since COVID-19 outbreak (March 2020). (A) Cumulative interactions, mutation features and binding regions annotated for SARS-CoV-2 (red), SARS-CoV1 (orange) and other *Coronaviridae* family members proteins (gray). Interactions include spoke-expanded binary relationships. Dots represent the date when the interaction was curated. (B) Amount of cumulative experimental evidence associated with unique binary pairs, captured over time, in each of the three datasets: SARS-CoV-1, SARS-CoV-2 and other *Coronaviridae* family members. Interactions include spoke-expanded binary relationships.

expression systems. Additionally, strong differences can arise during the selection of *bona fide* interactors, where multiple strategies can be used in order to clean up spurious detections. The lack of common interactions between the different studies focused on SARS-CoV-2 suggests that more of these systematic experiments are needed to increase reliability of biological conclusions extracted from this type of data.

Exploring the biological context of SARS-CoV-2 interactors suggests that the three HT studies are complementary. Pathway enrichment (Figure 4 and Supplementary Table 2) finds commonalities on expected pathways related with cell cycle, response to stress and infectious disease, and molecules synthesis and processing. Key pathways such as innate immune response or cytokine signaling are only found in the dataset of Li *et al.*, along with several intracellular signaling and metabolic routes. Only SARS-CoV-1 shows enrichment in lung tissue, but none of the sets reach the 5× enrichment threshold suggested by Jain and Tuteja (17) (Supplementary Figure 1, Supplementary Table 3). Finally, SARS-CoV-2 interactors from the three studies are found as components of exosome, deacetylases, ATPase transmembrane complexes and other mitochondrial complexes annotated in the Complex Portal (Supplementary Table 4) (30). We see an abundance of complexes involved in endocytosis in the study of Gordon *et al.*, protein production, Ca^{2+}-dependent cell signaling and cell cycle control are mostly represented in the dataset provided by Li *et al.* while proteins involved in transcriptional regulation and transport have been found in the interactome provided by Stukalov and collaborators.

Development of the IMEx coronavirus dataset complements other curation efforts that have been initiated in the light of the pandemic. Interaction resources BioGRID (31) and VirHostNet (the latter specialized in virus–host interaction data) (32) have produced shallow curation interaction datasets taken from COVID-19-related literature. The IMEx Coronavirus dataset has been used by the DisGeNET database (33) for contextual annotation with related disea ses (https://www.disgenet.org/downloads#section9). Also, members of the IMEx Consortium are involved in the COVID-19 Disease Map initiative (https://covid.pages.uni.lu/), where interaction information from the dataset is guiding COVID-19-related pathway curation. As an example, the list of PMIDs from IMEx Coronavirus dataset has

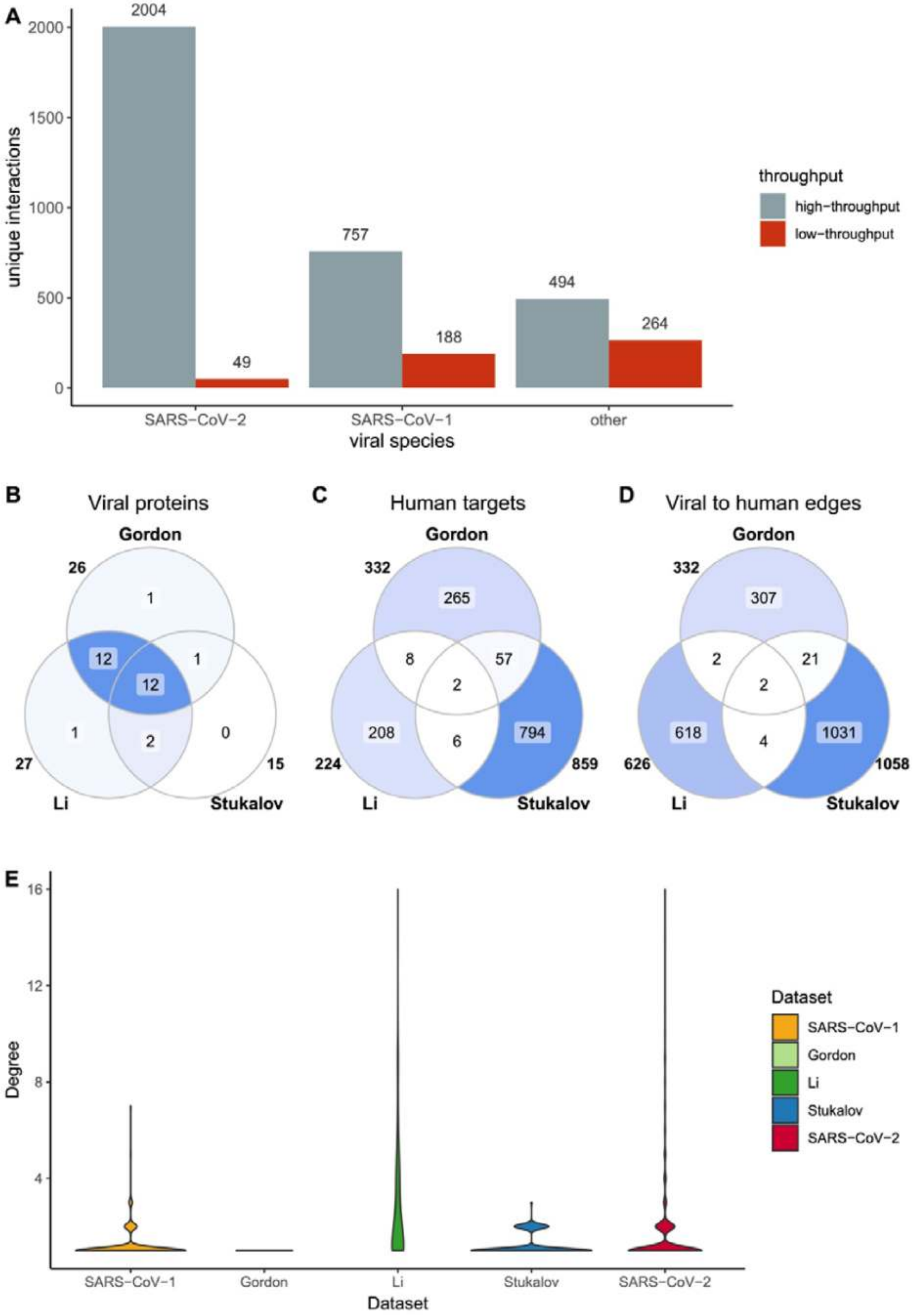

Figure 3. SARS-CoV-2 HT datasets comparison. SARS-CoV-2 datasets include Gordon *et al.*, Li *et al.* and Stukalov *et al.* (A) Distribution of HT/LT-derived interactions in the Coronavirus dataset. The number on top of each bar indicates the number of publications per category. HT publications are defined as those hosting more than 50 unique interacting pairs. (B-D) Overlap of viral proteins (B), human targets (C) and viral to human edges (D) across datasets. Numbers outside the circles indicate the size of the sets. (E) Distribution of number of interactions between viral proteins and human targets. In the y-scale, 1 indicates that one human protein interacts only with one viral protein, while 16 indicates that one human protein interacts with 16 viral proteins.

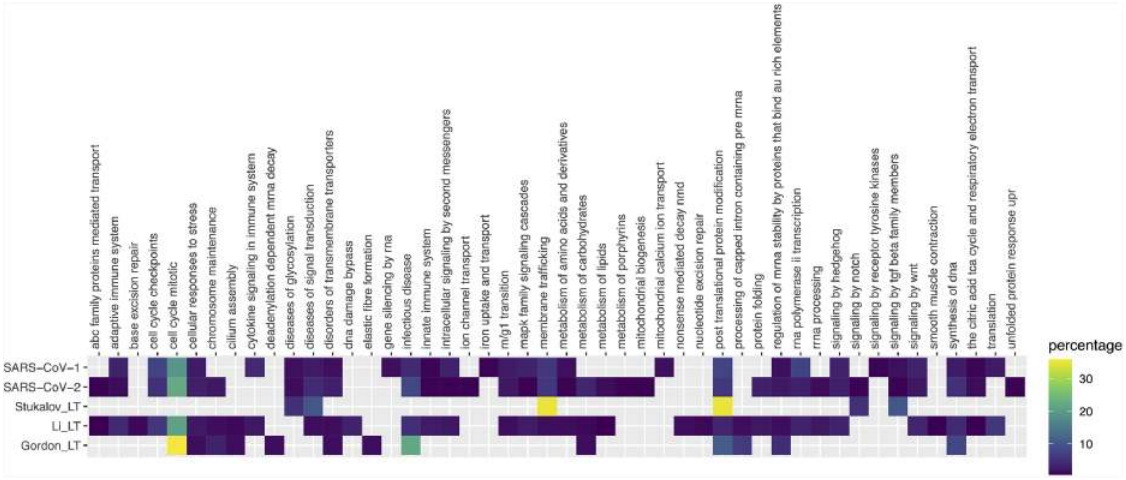

Figure 4. Pathway enrichment analysis of SARS-CoV-1, SARS-CoV-2 and Gordon_LT (Gordon plus low throughput), Li_LT (Li plus low throughput) and Stukalov_LT (Stukalov plus low throughput) datasets. Enrichment was performed using pathDIP (a database integrating 24 different pathway databases). Only human proteins were considered. The majority of enriched pathways were from Reactome database, so a mapping of each Reactome pathway to the parent pathway ontology was performed, and the heat map shows the percentage of pathways in each parent pathway over the total of pathways.

been used to screen papers containing causal interactions to build COVID-19 causal nobreaknetwork perturbed during SARS-COV2 infection by the SIGNOR 2.0 resource (34) (https://signor.uniroma2.it/covid/) and to select the Gordon *et al.* human interactors to integrate in the SIGNOR 2.0 network.

We are also involved in a parallel effort of curating protein complexes from SARS-CoV-1, SARS-CoV-2, MERS-CoV as well as their human target complexes in the Complex Portal (https://www.ebi.ac.uk/complexportal), linking to available experimental interaction evidence in IntAct when possible. This initiative is especially relevant because, as previously stated, coronaviruses increase the number of functional proteins produced by the viral genome by posttranslational cleavage of long polypeptide transcripts. The functionality and/or stability of these proteins are further increased through the formation of protein complexes, all of which have been catalogued into the Complex Portal. To date, 14 complexes have been identified in each strain, formed by viral–viral protein interactions, including homomeric assemblies such as the dimeric SARS-CoV-2 main protease complex (CPX-5685). Other reference entities such as the SARS-CoV-2 Spike–human ACE2 receptor complex (CPX-5683) have also been created to enable their identification in large omics datasets. All complexes have been annotated with Gene Ontology terms, describing their role in the virus lifecycle, which again will assist in the analysis of large omics-derived datasets.

Conclusions

Accurate and detailed representation of biological insight into public databases is a fundamental source of data for scientific discovery, even more so in a situation of accelerated research work such as the current pandemic. Our curation of molecular interactions related to *Coronaviridae* enables a systematic perspective of this data and greatly increases its interoperability. From our overview analysis of the data available so far, we can highlight how the SARS-CoV-2 data seem to be both biologically relevant, and thus informative for research on the disease, and strongly preliminary, so full consideration for its inherent incompleteness should be given when using it. The IMEx Consortium is expanding the resource as new data become available, striving to provide the most accurate possible picture of the *Coronaviridae* interactome.

Acknowledgements

The authors would like to thank Dr Yun Niu for assistance in pathway reduction analysis, Dr Serene Wong for useful discussion and Dr Rudolf Pillich for his help with NDEx networks upload and refinement.

Funding

The IntAct and Complex Portal databases and EMBL-EBI-based authors received funding from EMBL core funding, Open Targets (grant agreements OTAR-044 and OTAR02-048) and the Wellcome Trust grant INVAR (grant ref: 212925/Z/18/Z). IID and some computational analyses were in part supported by the grants from Ontario Research Fund (#34876), Natural Sciences Research Council (NSERC #203475), Canada Foundation for Innovation (CFI #29272, #225404, #33536), Buchan Foundation and IBM. Work by the UniProt database was supported by the National Eye Institute (NEI), National Human Genome Research Institute (NHGRI), National Heart, Lung, and Blood Institute (NHLBI), National Institute on Aging (NIA), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of General Medical Sciences (NIGMS), National Cancer Institute (NCI) and National Institute of Mental Health (NIMH) of the National Institutes of Health under award number (U24HG007822; the content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health). MatrixDB is supported by the Fondation pour la Recherche Medicale (DBI20141231336) and the Institut Français de Bioinformatique (Glycomatrix project, 2015). DIP database is supported by the National Institute of General Medical Sciences (R01GM123126). MINT is supported by ERC European Research Council Grant (322749 to G.C.), AIRC startup 21815 and AIRC Project IG 2017 (20322 to G.C.).

Supplementary data

Supplementary data are available at *Database* Online.

References

1. Rota, P.A., Oberste, M.S., Monroe, S.S. *et al.* (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, **300**, 1394–1399.
2. Wu, F., Zhao, S., Yu, B. *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature*, **579**, 265–269.
3. Ostaszewski, M., Mazein, A., Gillespie, M.E. *et al.* (2020) COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms. *Sci Data*, **7**, 136.
4. Orchard, S., Kerrien, S., Abbani, S. *et al.* (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.
5. Orchard, S., Ammari, M., Aranda, B. *et al.* (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
6. Licata, L., Briganti, L., Peluso, D. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
7. Xenarios, I., Salwínski, L., Duan, X.J. *et al.* (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.*, **30**, 303–305.
8. UniProt Consortium. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
9. Clerc, O., Deniaud, M., Vallet, S.D. *et al.* (2019) MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. *Nucleic Acids Res.*, **47**, D376–D381.
10. Kotlyar, M., Pastrello, C., Sheahan, N. *et al.* (2016) Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic Acids Res.*, **44**, D536–D541.
11. R Core Team. (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
12. Dowle, M. and Srinivasan, A. (2019) data.table: extension of 'data.frame'.
13. Gao, C.-H. (2019) ggVennDiagram: A "ggplot2" Implement of Venn diagram.
14. Ram, K. and Wickham, H. (2018) Wesanderson: a Wes Anderson Palette Generator.
15. Brown, K.R., Otasek, D., Ali, M. *et al.* (2009) NAViGaTOR: network analysis, visualization and graphing Toronto. *Bioinformatics*, **25**, 3327–3329.
16. Shannon, P., Markiel, A., Ozier, O. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
17. Jain, A. and Tuteja, G. (2019) TissueEnrich: tissue-specific gene enrichment analysis.*Bioinformatics*, **35**, 1966–1967.
18. Uhlén, M., Fagerberg, L., Hallström, B.M. *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
19. Rahmati, S., Abovsky, M., Pastrello, C. *et al.* (2020) pathDIP 4: an extended pathway annotations and enrichment analysis resource for human, model organisms and domesticated species. *Nucleic Acids Res.*, **48**, D479–D488.
20. Jassal, B., Matthews, L., Viteri, G. *et al.* (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **48**, D498–D503.
21. del-Toro, N., Dumousseau, M., Orchard, S. *et al.* (2013) A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res.*, **41**, W601–W606.
22. Calderone, A., Licata, L. and Cesareni, G. (2015) VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Res.*,**43**,D588–D592.
23. IMEx Consortium Curators Del-Toro, N., Duesbury, M., Koch, M. *et al.* (2019) Capturing variation impact on molecular interactions in the IMEx Consortium mutations data set. *Nat. Commun.*, **10**, 10.
24. Gordon, D.E., Jang, G.M., Bouhaddou, M. *et al.* (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, **583**, 459–468.
25. Stukalov, A., Girault, V., Grass, V. *et al.* (2020) Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV Systems Biology. https://doi.org/10.1101/2020.06.17.156455
26. Li, J., Guo, M., Tian, X. *et al.* (2020) Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med (N Y)*, 10.1016/j.medj.2020.07.002
27. Wodak, S.J., Vlasblom, J., Turinsky, A.L. *et al.* (2013) Protein-protein interaction networks: the puzzling riches. *Curr. Opin. Struct. Biol*, **23**, 941–953.
28. Venkatesan, K., Rual, J.-F., Vazquez, A. *et al.* (2009) An empirical framework for binary interactome mapping. *Nat. Methods*, **6**, 83–90.

29. Choi, S.G., Olivet, J., Cassonnet, P. *et al.* (2019) Maximizing binary interactome mapping with a minimal number of assays. *Nat. Commun.*, **10**, 3907.
30. Meldal, B.H.M., Bye-A-Jee, H., Gajdoš, L. *et al.* (2018) Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. *Nucleic Acids Res.*, 10.1093/nar/gky1001
31. Oughtred, R., Stark, C., Breitkreutz, B.-J. *et al.* (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Res.*, **47**, D529–D541.
32. Guirimand, T., Delmotte, S. and Navratil, V. (2015) VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res.*, **43**, D583–D587.
33. Piñero, J., Bravo, À., Queralt-Rosinach, N. *et al.* (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.*, **45**, D833–D839.
34. Licata, L., Lo Surdo, P., Iannuccelli, M. *et al.* (2020) SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. *Nucleic Acids Res.*, **48**, D504–D510.