Open science resources for the mass spectrometry-based analysis of SARS-CoV-2

Wout Bittremieux[1,2,*], **Charlotte Adams**[2], **Kris Laukens**[2], **Pieter C. Dorrestein**[1], **Nuno Bandeira**[1,3]

[1]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA; [2]Department of Computer Science, University of Antwerp, 2020 Antwerp, Belgium; [3]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA.

Corresponding author: wbittremieux@health.ucsd.edu

Abstract

The SARS-CoV-2 virus is the causative agent of the 2020 pandemic leading to the COVID-19 respiratory disease. With many scientific and humanitarian efforts ongoing to develop diagnostic tests, vaccines, and treatments for COVID-19, and to prevent the spread of SARS-CoV-2, mass spectrometry research, including proteomics, is playing a role in determining the biology of this viral infection. Proteomics studies are starting to lead to an understanding of the roles of viral and host proteins during SARS-CoV-2 infection, their protein-protein interactions, and post-translational modifications. This is beginning to provide insights into potential therapeutic targets or diagnostic strategies that can be used to reduce the long-term burden of the pandemic. However, the extraordinary situation caused by the global pandemic is also highlighting the need to improve mass spectrometry data and workflow sharing. We therefore describe freely available data and computational resources that can facilitate and assist the mass spectrometry-based analysis of SARS-CoV-2. We exemplify this by reanalyzing a virus-host interactome dataset to detect protein-protein interactions and identify host proteins that could potentially be used as targets for drug repurposing.

Keywords

SARS-CoV-2, mass spectrometry, open science, data sharing, virus-host interactome

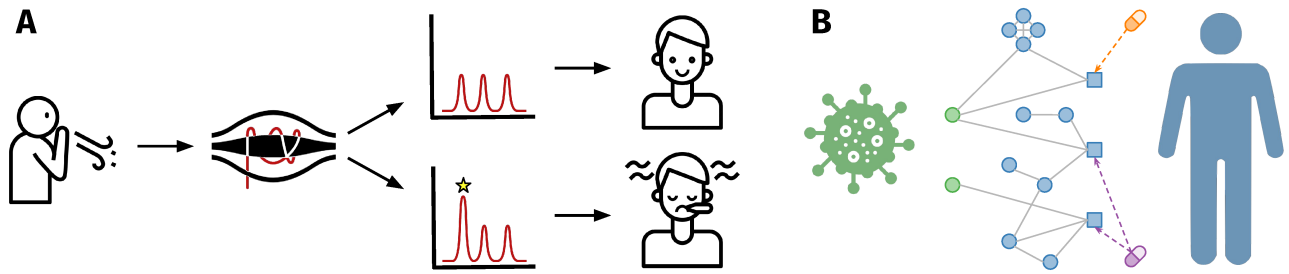

Figure 1: Mass spectrometry approaches to aid in combating SARS-CoV-2. **(A)** MS-based techniques can be used as a diagnostic test to determine whether people are infected and to assess disease severity. **(B)** Mass spectrometry can be used to understand the dynamics and mechanism of viral infection. Insights in changes in host cell protein dynamics and the virus-host interactome can be used to identify potential drug targets and develop antiviral therapies.

1 Introduction

An outbreak of the novel coronavirus SARS-CoV-2 in late 2019 proceeded to spread internationally in early 2020.[1] Leading to the COVID-19 respiratory disease, the SARS-CoV-2 virus presents a major societal challenge on a global scale. To date, COVID-19 has led to a hundred million infections and has resulted in millions of deaths worldwide, with these numbers still firmly increasing on a daily basis. Although the first COVID-19 vaccines have recently been approved,[2–4] until a sufficient number of people have been vaccinated to achieve herd immunity, governments worldwide have to continue to institute measures to mitigate spread of the virus, including quarantine and isolation procedures. Meanwhile, many researchers around the world have rallied to understand the spread and biology of viral infection by SARS-CoV-2 and to develop antiviral therapies. A detailed understanding of its fundamental infectious cycle is essential for the effective development of treatments to eliminate or control SARS-CoV-2. To this end, mass spectrometry (MS)-based research is playing a key role in gaining an understanding of the virus's biology and to support the development of diagnostics.[5]

Several studies propose to use mass spectrometry for diagnostic testing to detect SARS-CoV-2 infection as an alternative to reverse transcription polymerase chain reaction (RT-PCR) (figure 1A).[6–18] Often peptides from the abundant nucleocapsid protein are used as markers of SARS-CoV-2 infection. However, the reported efficacy of these studies varies wildly, from only 20 % sensitivity[18] up to near-perfect detection rates.[7,14,16] As such studies continue to emerge, it is important to provide essential details such as a comparison to the viral load detectability, the limit of detection, and the specificity when compared to other methods such as RT-PCR. Alternatively, studies that aim to detect SARS-CoV-2 infection indirectly by measuring changes in host proteins and metabolites are also being introduced.[19,20] However, as in this case the viral proteins are not directly and unambiguously detected, MS signature-based diagnostics are an indirect readout compared to RT-PCR or targeted MS methods. Additionally, several studies use long liquid chromatography gradients[6,10,11,15] or require advanced MS setups,[13,14] prohibiting the high throughput required for large-scale testing and rendering such methods, without major modifications, unfeasible in practice (for purposes other than research). Considering

these caveats, MS-based diagnostic approaches should currently be considered preliminary and extensive validation, which will likely take longer, is needed to confirm their robustness and accuracy.[21] This validation should include much larger sample sizes and performing the same experiment in different laboratories. Additionally, and this is especially relevant when employing an indirect readout of a signature as opposed to direct measurement of the viral proteins, a large-scale and heterogeneous population is required and disease assessments must be made to determine whether other diseases, such as related viral infections, do not confound the results. Alternatively, until such MS-based diagnostics can be assessed with tens to hundreds of thousands of samples that include a diversity of patients and a diverse set of diseases, and assuming their accuracy is satisfactory, they may be repurposed as a pre-screening method to triage patients for subsequent testing using unambiguous diagnostics that rely on a direct readout of the virus itself, such as sequencing (and perhaps in the future measurements of the viral proteins themselves by mass spectrometry).

Many MS-based SARS-CoV-2 studies employ underpowered and problematic statistical and machine learning strategies that risk overfitting and could lead to incorrect results. Some authors have correctly stated that their results should indeed only be considered preliminary,[6,9–11,13,15,18,19,22–26] whereas others have not clearly specified this caveat.[7,14,16,20,27] Furthermore, many of these studies do not make their raw data, identification results, feature tables, and code available (supplementary **??**), making it impossible to independently assess the validity of the results or to test alternative analysis workflows that could reveal additional data patterns that the original authors did not uncover with their chosen analysis approach. Of the 42 manuscripts we have evaluated that use mass spectrometry to study SARS-CoV-2, 22 studies have made their raw data fully available, five studies have additionally made their identification data or feature tables publicly available (18 more have made this data partially available), and only a single study has made its code publicly available (this was deemed not applicable for 15 studies, while two more studies have made their code partially available). Thus, a very large number of SARS-CoV-2 mass spectrometry papers are not adhering to the best practices for full data sharing and open science set by the community, funding bodies, and journals.[28–30] Perhaps in this case best practices can be skipped to some extent by authors, reviewers, and journals because there is an urgent need to get the work disseminated to address the SARS-CoV-2 pandemic. However, skipping best practices exacerbates the existing reproducibility crisis in science.[31] As current research on SARS-CoV-2 is highly publicized and scrutinized, if promising results later turn out to be irreproducible or incorrect, this might strongly damage the credibility of MS-based research. While there may be valid privacy or intellectual property reasons that raw data cannot be shared — and it needs to be disclosed when this is the case — downstream processed data tables and code are with very few exceptions shareable. "Data available upon request" should be considered unacceptable, especially as there are ample solutions available to share data and code, including specialized repositories for the deposition of MS data[32] and scientific sites to create permanently citable code archives.[33] Because different people have a different expertise and skills, data reanalyses might uncover novel knowledge that was overlooked in the original study. It is not only because when more eyes look at the data that it is possible to get a faster

and more complete understanding of its information content, but also that making data publicly available can spur the development of new computational methods that can lead to more accurate results. An example hereof is evolving data deposition requirements in the Protein Data Bank.[34] Whereas initially only processed data had to be deposited, with advances in algorithms the raw data could be processed directly to improve the quality of the protein structures, and thus not having the raw images from the X-ray available was a missed scientific opportunity. As a result, now the Protein Data Bank requires the deposition of raw image data as well. This is one of the reasons why we are highlighting the public MS data resources that are available (section 2), as such resources will help the larger community to fight the pandemic. Additionally, we also believe that research without its associated data and code available is much less impactful. Furthermore, there is reason to believe that, due to incorrect use of statistics and the irreproducible nature of the science itself, some of the reported results even go against the progress of fighting this pandemic.

Although mechanistic mass spectrometry-based studies are similarly beset by data and code sharing challenges, the mechanistic insights may ultimately lead to discoveries that improve the management, understanding of susceptibility patterns, prevention, and treatment of SARS-CoV-2 infection (figure 1B). Whereas the the initial SARS-CoV-2 genome was compiled using computational predictions and homology to other coronaviruses, novel research indicates that it likely contains several unannotated viral open reading frames.[35] MS techniques are excellently placed to confirm these results and help resolve ambiguities in the transcriptomic data.[36,37] Several groups have performed quantitative MS experiments to study the kinetics of viral infection over time, confirming that viral protein synthesis continuously increases after infection. Based on changes in host cell protein dynamics, activated pathways can be identified and used as potential drug targets.[38–40] Additionally, mass spectrometry can be used to study virus-host interactions. As viruses infect their hosts by a series of interactions between viral and host proteins,[41] the study of virus-host protein-protein interaction (PPI) networks is essential to understand the role of viral proteins during infection.[42] Several MS-based studies investigating the SARS-CoV-2 virus-host interactome have recently been released.[43–48] Knowledge on PPIs can be used to identify candidate druggable human proteins targeted by existing drugs.[43,48] Importantly, these studies are largely complementary as they use different experimental protocols to detect semi-overlapping sets of interacting proteins. This exemplifies how a fundamental understanding of virus-host interactions can be used to suggest possible repurposing of existing drugs and eventually develop novel therapeutics against SARS-CoV-2 infection.

Fortunately there are some commendable groups that continue to adhere to best scientific practices even in the middle of a pandemic. As of January 22, 2021, there are now 55 SARS-CoV-2 and related datasets accessible on MassIVE[49] via CoronaMassKB (section 2) and 36 datasets in the PRoteomics IDEntifications (PRIDE) database[50] for the community to reuse. Additionally, specialized SARS-CoV-2 data and computational infrastructures have been created to make the data more accessible, and we will describe how these resources can be used to study SARS-CoV-2. Besides a review of the resources and their utility, we also demonstrate how

Resource	Data type	Functionality
CoronaMassKB	proteomics, metabolomics	data repository, data reanalysis
COVID-19 Data Portal	genomics, proteomics	data indexing
COVID-19 UniProtKB	proteomics	protein sequence
Human Protein Atlas SARS-CoV-2	protein expression	data repository

Table 1: Overview of publicly available resources for the MS-based analysis of SARS-CoV-2.

publicly available proteomics data may be leveraged to drive discovery of PPIs. We provide a documented and reproducible example of how PPI data can be analyzed to determine virus-host interactions and show how this information can be potentially leveraged to suggest drugs that may be repurposed.

2 SARS-CoV-2 mass spectrometry public resources

An unprecedented scientific effort, with thousands of papers published on the topic in less than a year, is currently underway to study SARS-CoV-2 and try to develop antiviral therapies. Relevant scientific manuscripts, datasets, and novel insights are produced at a breakneck pace, making it difficult to keep up with the latest developments. To address this information disparity, several online resources which bundle knowledge at the data level on the MS-based analysis of SARS-CoV-2 in a few centralized locations and which facilitate and assist in the analysis of MS data have emerged (table 1):

CoronaMassKB is a community resource for sharing of MS data and analysis results built upon the MassIVE infrastructure.[49] It allows users to submit their coronavirus related data and (re)analysis results of existing datasets, and perform reanalyses directly using the online workflows that are available on MassIVE. It currently contains several reprocessed MS proteomics and metabolomics datasets and reanalysis results for SARS-CoV-2, SARS-CoV, MERS-CoV, and other coronaviruses.

COVID-19 Data Portal is a unified portal to access different types of omics data related to SARS-CoV-2, hosted by the European Bioinformatics Institute (EMBL-EBI). The data portal provides an easy access point to SARS-CoV-2 data that is available in various services hosted by EMBL-EBI, such as genomic sequences (from the European Nucleotide Archive[51]), expression data (including gene expression and single cell expression data from the Expression Atlas[52] and protein expression data hosted in the PRIDE database[50]), protein sequence data (from UniProtKB[53]), protein structures (from the Protein Data Bank in Europe[54]), and drug compound activity data (from ChEMBL[55]) and potential drug targets (complexes from the Complex Portal[56] and pathways from Reactome[57]).

COVID-19 UniProtKB provides the latest available pre-release UniProtKB[53] data for SARS-CoV-2 and

other entries related to the COVID-19 outbreak. It is updated as new relevant information becomes available, independent of the general UniProt release schedule. Note that the COVID-19 UniProtKB contains the full sequence of the SARS-CoV-2 replicase polyprotein 1ab as a single entry, rather than individual entries for the sixteen non-structural proteins into which it is auto-proteolytically processed.

Human Protein Atlas SARS-CoV-2 [58] contains tissue and cell expression patterns of known SARS-CoV-2 interacting human proteins, based on transcriptomics and antibody-based proteomics data that is curated from the scientific literature.

In addition to these specialized SARS-CoV-2 resources, relevant data has been deposited to the standard MS public data repositories. Several datasets are available through ProteomeXchange[59] in member repositories such as MassIVE,[49] PRIDE,[50] and iProx.[60] Additionally, several metabolomics coronavirus datasets are available on GNPS[61] and MetaboLights[62] as well. These datasets often form the starting point for (re)analyses included in the specialized SARS-CoV-2 data portals listed above. Additionally, PPI databases, such as IntAct[63] and other members of the International Molecular Exchange (IMEx) Consortium,[64] provide access to interaction information for SARS-CoV-2 that is being curated from the emerging scientific literature.[65]

When a novel virus, such as SARS-CoV-2, emerges, there is only a very limited amount of information available. Instead, molecular data for closely related, known, viruses can provide preliminary insights into viral activity.[66,67] Therefore, besides SARS-CoV-2 data, many of the resources described above also contain data for related coronaviruses, such as the SARS-CoV virus (causing the severe acute respiratory syndrome (SARS)) and the related MERS-CoV virus (causing the Middle East respiratory syndrome (MERS)). Genomic and proteomic information from these related coronaviruses can be used to investigate the functionality of homologous SARS-CoV-2 proteins and understand how host responses overlap or differ. For example, the receptor binding motif for Angiotensin I converting enzyme 2 (ACE2) binding by SARS-CoV, which it uses for host cell entry, is conserved in SARS-CoV-2, and SARS-CoV-2 similarly uses ACE2 for cellular entry.[68]

Besides these resources for data sharing, the COVID-19 Mass Spectrometry Coalition[69] is a collective and collaborative effort to share sample collection and processing protocols, and disseminate knowledge and best practices for MS analysis of COVID-19 samples. This is especially relevant because rigorous safety protocols need to be followed during analytical processing of COVID-19 samples to ensure that laboratory personnel does not get infected.

3 Public data sharing enables reanalysis of virus-host protein-protein interactions

To illustrate the power of public data sharing, we will demonstrate how existing PPI MS data can be reanalyzed using a different method to investigate the virus-host interactome and identify potential targets for drug repurposing. As PPIs drive most cellular functions, studying the virus-host interactome is essential to understand the molecular mechanisms of viral infection. Several high-throughput MS techniques can be used to determine protein interactions, such as affinity purification mass spectrometry (AP-MS),[70] proximity-dependent biotinylation (PDB) approaches,[71] and thermal protein profiling.[72] A widely used technique to identify protein interactions is AP-MS.[70] In an AP-MS experiment cells expressing a protein of interest (called the "bait") are cultured and harvested, allowing the purification of the protein by an affinity reagent targeting the protein itself or an epitope tag. The targeted protein is then isolated from nonspecific proteins through a series of washes and identified by mass spectrometry along with its interacting partners (called "preys"). Alternatively, during a proximity-dependent biotinylation experiment the bait is fused in-frame to a PDB enzyme, leading to the covalent biotinylation of nearby proteins. The advantage of PDB is that protein interactions or the integrity of organelles does not need to be maintained post-labeling because the covalently biotinylated preys can be captured using an affinity matrix.[71] These techniques have been instrumental in recent years to increase our understanding of protein interactions. As an example, the BioPlex project is a remarkable effort to map the human interactome.[73–75] Its latest iteration covers 70 % of the human proteome and includes 118 162 interactions among 14 586 proteins, resulting from a staggering 10 128 AP-MS experiments. This detailed knowledge on human protein interactions is highly valuable to determine normal cellular processes and analyze how viral infection hijacks cellular activity. With regard to SARS-CoV-2, several AP-MS and PDB-based studies investigating the virus-host interactome have recently been released.[43–48]

The outcome of a PPI experiment is a list of bait proteins and their potential interaction partners (preys). However, such unfiltered lists can contain a very high number of false positive interactions.[76] First, false positive interactions arise from incorrect protein identifications. Similar to standard spectrum identification, false positive identifications need to be controlled at the spectrum level, peptide level, and protein level.[77] Second, besides this general problem of false positive protein identifications, nonspecific interactors need to be removed.[76] Nonspecifically binding proteins, or background proteins, include highly abundant cellular proteins (e.g. tubulins and ribosomal proteins), proteins that bind to unfolded polypeptides (e.g. heat-shock proteins), and other contaminants. A further consideration when using MS-based techniques to determine PPIs is that, as opposed to binary interaction elucidation techniques, such as yeast two-hybrid (Y2H), they cannot differentiate between direct and indirect binding relationships. Consequently, using MS-based techniques it is not possible to determine whether two proteins directly interact with each other, or whether they interact indirectly as part

of a larger protein complex. In contrast, compared to Y2H and other *ex situ* binding assays, MS-based binding assays function *in situ*, i.e. they involve the binding of the protein of interest to host proteins inside of the cell. This decreases the detection of artificial interactions and modifications on the proteins and can be closer to the state found during infection.[78]

3.1 Filtering high-confidence interacting proteins

A crucial step in the computational analysis of PPI data is filtering out false positive interactions. The background is estimated over multiple unrelated MS runs in an experiment and proteins that are present at a higher abundance than the background are assumed to be interactors. Alternatively, the CRAPome is a repository of negative controls aggregated from multiple AP-MS experiments that can be used to help filter background contaminants in experiments conducted with similar protocols.[79] Several computational tools, such as CompPASS[80]/CompPASS-Plus,[73] SAINT[81]/SAINTexpress,[82] MiST,[83] and SFINX,[84] exist to filter nonspecific interactors and rank interacting proteins. These tools typically employ label-free protein quantification to score interactors and filter out low confidence interactions.

To demonstrate how publicly available proteomics data may be used to study the virus-host interactome and to illustrate the need for appropriate filtering of potential PPIs, we have reanalyzed the SARS-CoV-2 AP-MS data from Gordon et al. [43]. A fully end-to-end analysis, including detailed instructions and descriptions of the different steps, is available as a Jupyter notebook[85] at `https://github.com/bittremieux/sars_cov_2` (figure 2). Identification results have been deposited to the MassIVE repository with reanalysis identifier RPXD018240.3. First, raw files downloaded from PRIDE were converted to MGF files using ThermoRawFileParser (version 1.2.3).[86] Next, spectrum identification was performed using the Comet search engine (version 2016.01)[87] in the Crux toolkit (version 3.2)[88] against a concatenated FASTA file containing human protein sequences (Uniprot reviewed sequences downloaded on 2020/02/28),[89] the SARS-CoV-2 protein sequences, and the green fluorescent protein sequence (the AP-MS tag). The search settings included a concatenated target-decoy search, 10 ppm precursor mass tolerance, trypsin/P cleavage, removal of the precursor peak, and to only report the top scoring peptide per spectrum. Other search settings were kept at their default values, including a static carbamidomethylation modification of cysteine and no variable modifications. The peptide-spectrum matches (PSMs) were processed in Python 3.8 using Pyteomics (version 4.3.2)[90] to jointly filter them at a 1 % PSM and protein level false discovery rate (FDR), retaining 173 748 PSMs. Next, we have used SAINTexpress[82] to score PPIs based on spectral counts. In the notebook, we describe the format of the SAINTexpress input files and demonstrate how to generate them from the PSM data. Prior to SAINTexpress filtering there were 44 992 unique bait–prey combinations that form potential interacting proteins. After filtering the SAINTexpress results at 1 % FDR and requiring at least 20 spectra per bait–prey pair, 412 PPIs were retained. This 100-fold

□

Figure 2: Example Jupyter notebook to illustrate the benefits of open science for the analysis of SARS-CoV-2 data. The notebook contains a step-by-step reanalysis of the AP-MS data from Gordon et al. [43], including spectrum identification, PPI filtering and network visualization, and identifying potential targets for drug repurposing. The full notebook can be found at `https://github.com/bittremieux/sars_cov_2`.

reduction in PPIs clearly illustrates the need for stringent filtering to remove spurious interactions.

In their original analysis, Gordon et al. [43] report 332 high-confidence interactions. The difference in detected PPIs can largely be attributed to a difference in filtering strategies. Whereas Gordon et al. [43] have used a two-step combination of SAINTexpress[82] and MiST[83] filtering with different thresholds to identify high-confidence PPIs, here we have filtered on the FDR reported by SAINTexpress and a minimum spectra requirement. We observe that 319 of the 412 PPIs obtained from the reanalysis are present in the unfiltered data by Gordon et al. [43]. Additionally, 93 new PPIs are detected, which can be attributed to the fact that different search engines were used for spectrum identification. For the 93 novel PPIs the human protein targets of each SARS-CoV-2 bait were tested for enrichment of gene ontology (GO) biological process terms using g:Profiler (version e102_eg49_p15_7a9b4d6).[91] GO terms with a p value below 0.01 (after g:SCS multiple testing correction) were considered significant (supplementary **??**).

Several databases exist that collect information on PPIs,[92] such as IntAct,[63] MINT,[93] BioGrid,[94] STRING,[95] etc. These databases collect molecular interaction data that is derived and curated from the literature or that is predicted from genomic context information. Several of these data repositories are members of the IMEx Consortium,[64] which has established a single set of rules and guidelines for capturing PPI data and provides a unified interface for computational access to these data resources.[96] The IMEx Consortium has recently released a curated dataset of molecular interactions for SARS-CoV-2 and other coronaviruses, containing over 7000 binarized interactions.[65] We demonstrate in the notebook how the PSICQUIC web service[96] can be used to automatically query the curated interactions from IMEx. The comparison with the confirmed interactions (figure 3) indicates that reprocessing existing datasets with alternative computational tools recovers known, high-quality interactions, but can also suggest new interactions. This clearly demonstrates the value of public data sharing, as reanalysis can provide different perspectives on the same published data.

3.2 Virus-host protein interactions provide targets for drug repurposing

Drug repurposing is an especially attractive strategy in the early phase of an epidemic to provide immediate treatment.[98,99] The advantage of repurposing existing drugs is that no expensive and time-consuming drug development cycle has to be completed but that the drugs are readily available. Additionally, existing drugs have known safety profiles and are already familiar to clinicians working with these drugs. As an example,

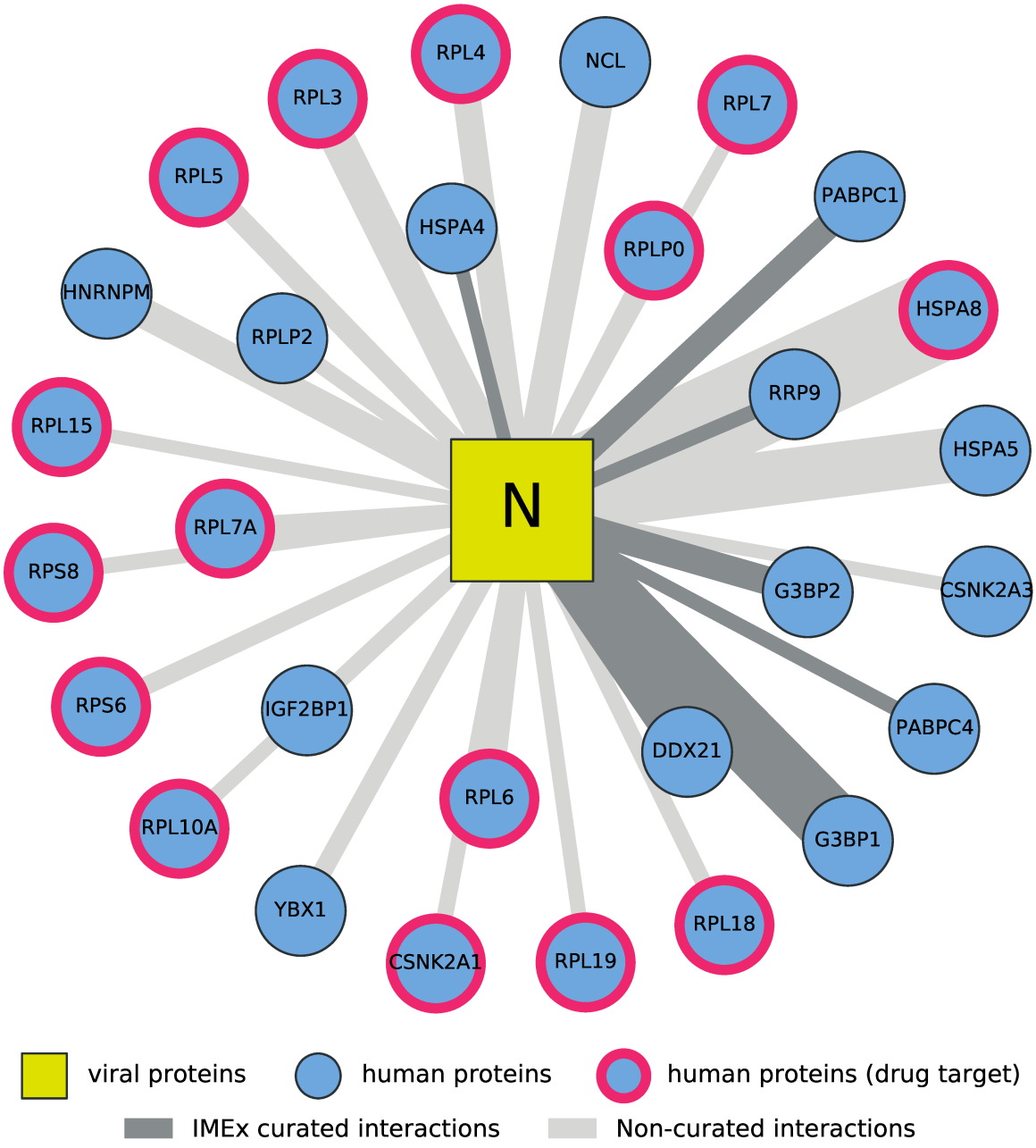

Figure 3: PPI network for the SARS-CoV-2 nucleocapsid (N) protein, visualized using NetworkX (version 2.4).[97] The weight of the edges is proportional to the average spectral counts of the preys across the biological replicates of the bait. Edges for interactions that have been manually extracted from the literature by IMEx curators[65] are colored dark gray, whereas edges for potentially novel interactions are colored light gray. Prey (human) proteins that are a target for existing drugs retrieved from ChEMBL[55] are indicated with a pink border.

dexamethasone, a commonly used corticosteroid, has been shown to reduce mortality among severe COVID-19 patients receiving invasive mechanical ventilation.[100]

Conventional antiviral drug therapies mainly target the virus itself. Most coronaviruses encode a large surface protein, the spike protein, which is responsible for receptor binding and membrane fusion. Antibodies that bind the spike protein prevent its attachment to the host cell and can potentially neutralize viral infection.[101] As such, although to date no vaccine or specific antiviral drug has been approved for the previous coronavirus SARS-CoV, its spike protein has been identified as a potential target for vaccine and therapeutic development.[102] Similarly, the spike protein was quickly identified as an antigenic target for development of a vaccine against SARS-CoV-2 as well.[101] Alternatively, instead of targeting viral proteins, knowledge about the virus-host interactome can be used to identify potentially relevant host protein targets for antiviral therapy.[103,104] For instance, Hoffmann et al. [68] show that the cellular serine protease TMPRSS2 primes the SARS-CoV-2 spike protein for entry, and that the serine protease inhibitor camostat mesylate, which is active against TMPRSS2, blocks SARS-CoV-2 infection of lung cells. Additionally, by matching PPI information determined via AP-MS to chemical databases, Gordon et al. [43] were able to identify 67 human proteins targeted by existing drugs, including drugs approved by the Food and Drug Administration (FDA), compounds in clinical trials, and preclinical compounds.

Similarly, the supplementary Jupyter notebook demonstrates how cheminformatics approaches can be used to find existing drugs that target host proteins that interact with SARS-CoV-2 proteins in the reprocessed AP-MS data from Gordon et al. [43]. Several databases of bioactive molecules, their chemical properties, and drug target information exist, including ChEMBL,[55] DrugBank,[105] the Therapeutic Target Database,[106] and BindingDB.[107] We have used the ChEMBL Python client library (version 0.10.1)[108] to automatically retrieve FDA-approved drug compounds that target the human proteins that were found to interact with SARS-CoV-2 (figure 3). This results in 2982 drug–protein pairs consisting of 51 unique human proteins that are targeted by 357 unique drugs. Interestingly, five drugs are currently being evaluated in clinical trials registered at ClinicalTrials.gov (supplementary **??**). These are (i) colchicine, an anti-inflammatory drug which has been shown to improve time to clinical deterioration;[109] (ii) cd24fc, which has immune checkpoint inhibitory and anti-inflammatory activities and which was originally developed for the treatment of graft-versus-host disease; (iii) selinexor, an anticancer drug that blocks the cellular protein XPO1, which mediates the export of viral proteins out of the nucleus,[110] (iv) silmitasertib, an anticancer drug that is an inhibitor of casein kinase II, whose activity has been shown to be strongly upregulated after SARS-CoV-2 infection,[39] and (v) paclitaxel, an anticancer drug. This demonstrates how the study of virus-host interactions can lead to mechanistic insights into viral infection to identify host proteins that can be used as potential targets for drug repurposing.

Furthermore, multiple approaches to evaluate drug targets *in silico* are being pursued.[111–113] As an example, based on the BenevolentAI knowledge graph,[114] which consists of medical information extracted from the

scientific literature using machine learning, the drug baricitinib has been predicted to reduce the ability of SARS-CoV-2 to infect lung cells,[112,115] which effect was recently confirmed in a clinical trial.[116] However, baricitinib also blocks JAK-STAT signaling, which is thought to lead to an increased risk of herpes zoster and simplex infection. As such, the use of baricitinib should currently be considered with caution.[117,118] As the scientific and medical community gains a deeper understanding into the dynamics of SARS-CoV-2 infection, additional drug repurposing targets might be identified. However, these recommendations should initially only be considered as tentative treatment options, which have to be carefully validated in randomized controlled trials.

3.3 The underexplored role of post-translational modifications in SARS-CoV-2 infection

Although viral protein complexity is often limited by its relatively small genome, post-translational modifications (PTMs) offer additional functional diversity within the limited genetic space. PTMs play an important role in enhancing the multifunctional nature of viral proteins by contributing to both cellular responses to infection and viral hijacking of the host.[119] Although our current knowledge of the presence and role of PTMs in SARS-CoV-2 is still limited, viral replication and pathogenesis of other coronaviruses has been shown to be influenced by the presence of PTMs.[120] When the SARS-CoV spike glycoprotein was treated with PNGase F to remove N-linked glycans it was no longer recognized by neutralizing antisera raised against purified virions,[121] suggesting that N-linked glycosylation may play an important role in constituting the native structure of the spike protein and thereby affecting its antigenicity. The SARS-CoV nucleocapsid protein can be phosphorylated by multiple host kinases. Inhibition of the host protein glycogen synthase kinase-3 significantly reduced the phosphorylation level of the nucleocapsid protein and led to a decrease in viral titer and cytopathic effects.[122] Additionally, coronaviruses can interfere with host PTMs as well. The replicase polyprotein encoded by coronavirus gene 1 is processed into sixteen nonstructural proteins by two viral proteases, papain-like protease (PLpro) and 3C-like protease.[123] Apart from its protease activity, SARS-CoV PLpro,[124] MERS-CoV PLpro,[125] as well as PLpro of other coronaviruses have also been shown to possess deubiquitinating activity.

Both viral and host PTMs might similarly play a role in infection by SARS-CoV-2, and mass spectrometry is being used to provide insights into the modification profile of SARS-CoV-2 proteins. There is a high degree of structural similarity and sequence conservation between the SARS-CoV spike protein and the SARS-CoV-2 spike protein.[126] As the currently approved vaccines and other vaccine candidates are focused on the spike protein,[101] detailed knowledge of its glycosylation status is highly relevant.[127–131] Additionally, numerous phosphorylation sites have been reported within critical SARS-CoV-2 proteins,[36,39,40,48,132] with extensive temporal changes in phosphorylation of host and viral proteins over the time course of SARS-CoV-2 infection.[40] Similarly to the SARS-CoV nucleocapsid protein, the SARS-CoV-2 nucleocapsid protein is phosphorylated by host

kinases, leading to the hypothesis that this phosphorylation may modulate nucleocapsid protein function and influence cytopathic effect.[39] Furthermore, although the functional importance of the SARS-CoV-2 phosphoproteins is as of yet unknown, knowledge of viral PTMs can provide new targets for the development of antiviral treatments.[39,48]

Decoding the role of these and other PTMs can provide essential insights into viral infection and aid in developing effective antiviral treatments. Unfortunately, the computational identification of PTMs in MS data is often not straightforward.[133] In the standard MS identification paradigm each PTM to be considered has to be explicitly specified in the search settings. However, each additional PTM that is included leads to an increase in search space, resulting in an increased computational load and decreased sensitivity and specificity. As a result, most proteomics studies only consider a limited number of the most prevalent PTMs (mostly sample-handling artifacts[134]). In contrast, open modification searching (OMS) is an increasingly popular strategy to identify any type of modification in an unbiased fashion.[135] Rather than using a small precursor mass window as in a standard search, OMS works by allowing a modified spectrum to match against its unmodified peptide by using a very wide precursor mass window, exceeding the delta mass induced by any PTMs that might be present. Subsequently the presence and types of modifications can be derived from the observed precursor mass difference. Historically OMS suffered from large computational requirements due to the use of a very wide precursor mass window, precluding the unbiased profiling of modifications at a proteome-wide scale. In contrast, novel OMS tools use advanced algorithmic approaches to efficiently process this very large search space, such as using a fragment ion index,[136] sequence tags,[137,138] or nearest neighbor searching[139,140] to filter the search space; or using multinotch searching to look for specific PTMs.[141] Using these advanced OMS tools presents a highly interesting data reanalysis opportunity for the unbiased investigation into the role of PTMs during viral infection by SARS-CoV-2.

4 Conclusion

While the current SARS-CoV-2 pandemic presents an unprecedented global challenge, the scientific community has likewise responded with extraordinary speed and focus. With many researchers pivoting towards the analysis of SARS-CoV-2, a massive amount of novel scientific knowledge is currently being generated on a daily basis, and it can be a daunting task to keep up with this rapidly growing body of research. We have highlighted a few centralized data portals that contain resources to facilitate the MS-based analysis of SARS-CoV-2. These data portals play an important role in featuring relevant scientific work and efficiently sharing emerging insights. Despite the urgency imposed by the pandemic we would like to encourage the scientific community to keep adhering to best practices to ensure the highest possible quality of research. Importantly, this includes public data and code sharing, ideally using the aforementioned resources, to avoid exacerbating the reproducibility

crisis, impart confidence in the published results, and enable and facilitate public data reanalyses.

As a brief demonstration of the potential of public data sharing we have reanalyzed a recently published SARS-CoV-2 PPI dataset. We have shown how known and novel protein interactions can be found and how this information can be used to identify candidates for drug repurposing. Additionally, we have highlighted opportunities for further reanalyses of public data to investigate the unknown role of protein modifications during viral infection. Crucially, such reanalyses to explore the virus-host interactome in deeper detail are only possible if the data is publicly accessible. Additional experimental studies remain necessary as well to obtain a full understanding of viral infection. For example, complementary experimental techniques, including AP-MS and BioID, are required to fully map the SARS-CoV-2 virus-host interactome.[43–48]

These are challenging times for a lot of people, both in their personal lives as well as in their professional lives. Nevertheless, seeing the collegiality in the scientific community, by collaborating on projects to understand SARS-CoV-2 and by emotionally supporting friends and colleagues in these taxing circumstances, makes us feel optimistic that together we will overcome these challenges.

Acknowledgments

W.B. is a postdoctoral researcher of the Research Foundation – Flanders (grant 12W0418N).

Conflict of interest

P.C.D. is on the scientific advisory board of Sirenas, Cybele Microbiome, and Galileo.

Supporting information

The following supporting information is available free of charge at ACS website https://pubs.acs.org/: Supplementary Table S1. Data and code availability for MS-based studies of SARS-CoV-2; Supplementary Table S2. Gene ontology enrichment results for newly detected virus–host interactions; Supplementary Table S3. COVID-19 drug repurposing candidates evaluated in clinical trials.

References

(1) Wu, F., Zhao, S., Yu, B., Chen, Y.-M., et al. A New Coronavirus Associated with Human Respiratory Disease in China.*Nature* **2020**, *579*, 265–269, DOI: 10.1038/s41586-020-2008-3.

(2) Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine.*New England Journal of Medicine* **2020**, DOI: 10.1056/NEJMoa2035389.

(3) Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.*New England Journal of Medicine* **2020**, *383*, 2603–2615, DOI: 10.1056/NEJMoa2034577.

(4) Voysey, M., Clemens, S. A. C., Madhi, S. A., Weckx, L. Y., et al. Safety and Efficacy of the ChAdOx1 nCoV-19 Vaccine (AZD1222) against SARS-CoV-2: An Interim Analysis of Four Randomised Controlled Trials in Brazil, South Africa, and the UK.*The Lancet* **2020**, *397*, 99–111, DOI: 10.1016/S0140-6736(20)32661-1.

(5) Grenga, L., Armengaud, J. Proteomics in the COVID-19 Battlefield: First Semester Check-up.*PROTEOMICS* **2020**, *21*, 2000198, DOI: 10.1002/pmic.202000198.

(6) Bezstarosti, K., Lamers, M. M., Haagmans, B. L., Demmers, J. A. Targeted Proteomics for the Detection of SARS-CoV-2 Proteins.*bioRxiv* **2020**, DOI: 10.1101/2020.04.23.057810.

(7) Cardozo, K. H. M., Lebkuchen, A., Okai, G. G., Schuch, R. A., et al. Fast and Low-Cost Detection of SARS-CoV-2 Peptides by Tandem Mass Spectrometry in Clinical Samples.*Research Square* **2020**, DOI: 10.21203/rs.3.rs-28883/v1.

(8) Cazares, L. H., Chaerkady, R., Samuel Weng, S. H., Boo, C. C., et al. Development of a Parallel Reaction Monitoring Mass Spectrometry Assay for the Detection of SARS-CoV-2 Spike Glycoprotein and Nucleoprotein.*Analytical Chemistry* **2020**, *92*, 13813–13821, DOI: 10.1021/acs.analchem.0c02288.

(9) Gouveia, D., Grenga, L., Gaillard, J., Gallais, F., et al. Shortlisting SARS-CoV-2 Peptides for Targeted Studies from Experimental Data-dependent Acquisition Tandem Mass Spectrometry Data.*PROTEOMICS* **2020**, *20*, 2000107, DOI: 10.1002/pmic.202000107.

(10) Gouveia, D., Miotello, G., Gallais, F., Gaillard, J.-C., et al. Proteotyping SARS-CoV-2 Virus from Nasopharyngeal Swabs: A Proof-of-Concept Focused on a 3 Min Mass Spectrometry Window.*Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00535.

(11) Ihling, C., Tänzler, D., Hagemann, S., Kehlen, A., et al. Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients.*Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00280.

(12) Iles, R. K., Zmuidinaite, R., Iles, J. K., Carnell, G., et al. Development of a Clinical MALDI-ToF Mass Spectrometry Assay for SARS-CoV-2: Rational Design and Multi-Disciplinary Team Work.*Diagnostics* **2020**, *10*, 746, DOI: 10.3390/diagnostics10100746.

(13) Nikolaev, E. N., Indeykina, M. I., Brzhozovskiy, A. G., Bugrova, A. E., et al. Mass-Spectrometric Detection of SARS-CoV-2 Virus in Scrapings of the Epithelium of the Nasopharynx of Infected Patients via Nucleocapsid N Protein.*Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00412.

(14) Renuse, S., Vanderboom, P. M., Maus, A. D., Kemp, J. V., et al. Development of Mass Spectrometry-Based Targeted Assay for Direct Detection of Novel SARS-CoV-2 Coronavirus from Clinical Specimens.*bioRxiv* **2020**, DOI: 10.1101/2020.08.05.20168948.

(15) Rivera, B., Leyva, A., Portela, M. M., Moratorio, G., et al. Quantitative Proteomic Dataset from Oro- and Naso-Pharyngeal Swabs Used for COVID-19 Diagnosis: Detection of Viral Proteins and Host's Biological Processes Altered by the Infection.*Data in Brief* **2020**, *32*, 106121, DOI: 10.1016/j.dib.2020.106121.

(16) Singh, P., Chakraborty, R., Marwal, R., Radhakrishan, V. S., et al. A Rapid and Sensitive Method to Detect SARS-CoV-2 Virus Using Targeted-Mass Spectrometry.*Journal of Proteins and Proteomics* **2020**, *11*, 159–165, DOI: 10.1007/s42485-020-00044-9.

(17) Whetton, A. D., Preston, G. W., Abubeker, S., Geifman, N. Proteomics and Informatics for Understanding Phases and Identifying Biomarkers in COVID-19 Disease.*Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00326.

(18) Zecha, J., Lee, C.-Y., Bayer, F. P., Meng, C., et al. Data, Reagents, Assays and Merits of Proteomics for SARS-CoV-2 Research and Testing.*Molecular & Cellular Proteomics* **2020**, *19*, mcp.RA120.002164, DOI: 10.1074/mcp.RA120.002164.

(19) Kimhofer, T., Lodge, S., Whiley, L., Gray, N., et al. Integrative Modeling of Quantitative Plasma Lipoprotein, Metabolic, and Amino Acid Data Reveals a Multiorgan Pathological Signature of SARS-CoV-2 Infection.*Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00519.

(20) Nachtigall, F. M., Pereira, A., Trofymchuk, O. S., Santos, L. S. Detection of SARS-CoV-2 in Nasal Swabs Using MALDI-MS.*Nature Biotechnology* **2020**, *38*, 1168–1173, DOI: 10.1038/s41587-020-0644-7.

(21) Grossegesse, M., Hartkopf, F., Nitsche, A., Schaade, L., et al. Perspective on Proteomics for Virus Detection in Clinical Samples.*Journal of Proteome Research* **2020**, *19*, 4380–4388, DOI: 10.1021/acs.jproteome.0c00674.

(22) D'Alessandro, A., Thomas, T., Dzieciatkowska, M., Hill, R. C., et al. Serum Proteomics in COVID-19 Patients: Altered Coagulation and Complement Status as a Function of IL-6 Level.*Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00365.

(23) Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., et al. Large-Scale Multi-Omic Analysis of COVID-19 Severity.*Cell Systems* **2021**, *12*, 23–40.e7, DOI: 10.1016/j.cels.2020.10.003.

(24) Shen, B., Yi, X., Sun, Y., Bi, X., et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera.*Cell* **2020**, *182*, S0092867420306279, DOI: 10.1016/j.cell.2020.05.032.

(25) Song, J.-W., Lam, S. M., Fan, X., Cao, W.-J., et al. Omics-Driven Systems Interrogation of Metabolic Dysregulation in COVID-19 Pathogenesis. *Cell Metabolism* **2020**, *32*, DOI: 10.1016/j.cmet.2020.06.016.

(26) Wendt, R., Kalbitz, S., Lübbert, C., Kellner, N., et al. Urinary Proteomics Associates with COVID-19 Severity: Pilot Proof-of-Principle Data and Design of a Multicentric Diagnostic Study. *PROTEOMICS* **2020**, DOI: 10.1002/pmic.202000202.

(27) Messner, C. B., Demichev, V., Wendisch, D., Michalick, L., et al. Clinical Classifiers of COVID-19 Infection from Novel Ultra-High-Throughput Proteomics. *bioRxiv* **2020**, DOI: 10.1101/2020.04.27.20081810.

(28) Journal of Proteome Research – Author Guidelines, https://publish.acs.org/publish/author_guidelines?coden=jprobs, 2020.

(29) Molecular & Cellular Proteomics – Required manuscript content and publication guidelines for Molecular & Cellular Proteomics, https://www.mcponline.org/page/content/mass-spec-guidelines, 2020.

(30) Nature Research – Reporting standards and availability of data, materials, code and protocols, https://www.nature.com/nature-research/editorial-policies/reporting-standards, 2020.

(31) Baker, M. 1,500 Scientists Lift the Lid on Reproducibility. *Nature* **2016**, *533*, 452–454, DOI: 10.1038/533452a.

(32) Vizcaíno, J. A., Deutsch, E. W., Wang, R., Csordas, A., et al. ProteomeXchange Provides Globally Coordinated Proteomics Data Submission and Dissemination. *Nature Biotechnology* **2014**, *32*, 223–226, DOI: 10.1038/nbt.2839.

(33) Perkel, J. Democratic Databases: Science on GitHub. *Nature* **2016**, *538*, 127–128, DOI: 10.1038/538127a.

(34) Berman, H., Henrick, K., Nakamura, H. Announcing the Worldwide Protein Data Bank. *Nature Structural & Molecular Biology* **2003**, *10*, 980–980, DOI: 10.1038/nsb1203-980.

(35) Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., et al. The Coding Capacity of SARS-CoV-2. *Nature* **2020**, DOI: 10.1038/s41586-020-2739-1.

(36) Davidson, A. D., Williamson, M. K., Lewis, S., Shoemark, D., et al. Characterisation of the Transcriptome and Proteome of SARS-CoV-2 Using Direct RNA Sequencing and Tandem Mass Spectrometry Reveals Evidence for a Cell Passage Induced in-Frame Deletion in the Spike Glycoprotein That Removes the Furin-like Cleavage Site. *Genome Medicine* **2020**, *12*, 68, DOI: 10.1186/s13073-020-00763-0.

(37) Weingarten-Gabbay, S., Klaeger, S., Sarkizova, S., Pearlman, L. R., et al. SARS-CoV-2 Infected Cells Present HLA-I Peptides from Canonical and out-of-Frame ORFs. *bioRxiv* **2020**, DOI: 10.1101/2020.10.02.324145.

(38) Bojkova, D., Klann, K., Koch, B., Widera, M., et al. Proteomics of SARS-CoV-2-Infected Host Cells Reveals Therapy Targets. *Nature* **2020**, *583*, 469–472, DOI: 10.1038/s41586-020-2332-7.

(39) Bouhaddou, M., Memon, D., Meyer, B., White, K. M., et al. The Global Phosphorylation Landscape of SARS-CoV-2 Infection.*Cell* **2020**, *182*, P685–712.e19, DOI: 10.1016/j.cell.2020.06.034.

(40) Klann, K., Bojkova, D., Tascher, G., Ciesek, S., et al. Growth Factor Receptor Signaling Inhibition Prevents SARS-CoV-2 Replication.*Molecular Cell* **2020**, *80*, P164–174.E4, DOI: 10.1016/j.molcel.2020.08.006.

(41) Brito, A. F., Pinney, J. W. Protein–Protein Interactions in Virus–Host Systems.*Frontiers in Microbiology* **2017**, *8*, DOI: 10.3389/fmicb.2017.01557.

(42) Perrin-Cocon, L., Diaz, O., Jacquemin, C., Barthel, V., et al. The Current Landscape of Coronavirus-Host Protein–Protein Interactions.*Journal of Translational Medicine* **2020**, *18*, DOI: 10.1186/s12967-020-02480-z.

(43) Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., et al. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing.*Nature* **2020**, *583*, 459–468, DOI: 10.1038/s41586-020-2286-9.

(44) Laurent, E. M., Sofianatos, Y., Komarova, A., Gimeno, J.-P., et al. Global BioID-Based SARS-CoV-2 Proteins Proximal Interactome Unveils Novel Ties between Viral Polypeptides and Host Factors Involved in Multiple COVID19-Associated Mechanisms.*bioRxiv* **2020**, DOI: 10.1101/2020.08.28.272955.

(45) Li, J., Guo, M., Tian, X., Wang, X., et al. Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis.*Med* **2020**, *1*, 1–15, DOI: 10.1016/j.medj.2020.07.002.

(46) Samavarchi-Tehrani, P., Abdouni, H., Knight, J. D., Astori, A., et al. A SARS-CoV-2 – Host Proximity Interactome.*bioRxiv* **2020**, DOI: 10.1101/2020.09.03.282103.

(47) St-Germain, J. R., Astori, A., Samavarchi-Tehrani, P., Abdouni, H., et al. A SARS-CoV-2 BioID-Based Virus-Host Membrane Protein Interactome and Virus Peptide Compendium: New Proteomics Resources for COVID-19 Research.*bioRxiv* **2020**, DOI: 10.1101/2020.08.28.269175.

(48) Stukalov, A., Girault, V., Grass, V., Bergant, V., et al. Multi-Level Proteomics Reveals Host-Perturbation Strategies of SARS-CoV-2 and SARS-CoV.*bioRxiv* **2020**, DOI: 10.1101/2020.06.17.156455.

(49) Wang, M., Wang, J., Carver, J., Pullman, B. S., et al. Assembling the Community-Scale Discoverable Human Proteome.*Cell Systems* **2018**, *7*, 412–421.e5, DOI: 10.1016/j.cels.2018.08.004.

(50) Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., et al. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data.*Nucleic Acids Research* **2019**, *47*, D442–D450, DOI: 10.1093/nar/gky1106.

(51) Amid, C., Alako, B. T. F., Balavenkataraman Kadhirvelu, V., Burdett, T., et al. The European Nucleotide Archive in 2019.*Nucleic Acids Research* **2019**, DOI: 10.1093/nar/gkz1063.

(52) Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A. M.-P., et al. Expression Atlas Update: From Tissues to Single Cells.*Nucleic Acids Research* **2019**, DOI: 10.1093/nar/gkz947.

(53) The UniProt Consortium UniProt: A Hub for Protein Information.*Nucleic Acids Research* **2015**, *43*, D204–D212, DOI: 10.1093/nar/gku989.

(54) Mir, S., Alhroub, Y., Anyango, S., Armstrong, D. R., et al. PDBe: Towards Reusable Data Delivery Infrastructure at Protein Data Bank in Europe.*Nucleic Acids Research* **2018**, *46*, D486–D492, DOI: 10.1093/nar/gkx1070.

(55) Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., et al. ChEMBL: Towards Direct Deposition of Bioassay Data.*Nucleic Acids Research* **2019**, *47*, D930–D940, DOI: 10.1093/nar/gky1075.

(56) Meldal, B. H. M., Bye-A-Jee, H., Gajdoš, L., Hammerová, Z., et al. Complex Portal 2018: Extended Content and Enhanced Visualization Tools for Macromolecular Complexes.*Nucleic Acids Research* **2019**, *47*, D550–D558, DOI: 10.1093/nar/gky1001.

(57) Jassal, B., Matthews, L., Viteri, G., Gong, C., et al. The Reactome Pathway Knowledgebase.*Nucleic Acids Research* **2019**, DOI: 10.1093/nar/gkz1031.

(58) Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., et al. Tissue-Based Map of the Human Proteome.*Science* **2015**, *347*, 1260419–1260419, DOI: 10.1126/science.1260419.

(59) Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., et al. The ProteomeXchange Consortium in 2017: Supporting the Cultural Change in Proteomics Public Data Deposition.*Nucleic Acids Research* **2017**, *45*, D1100–D1106, DOI: 10.1093/nar/gkw936.

(60) Ma, J., Chen, T., Wu, S., Yang, C., et al. iProX: An Integrated Proteome Resource.*Nucleic Acids Research* **2019**, *47*, D1211–D1217, DOI: 10.1093/nar/gky869.

(61) Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., et al. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking.*Nature Biotechnology* **2016**, *34*, 828–837, DOI: 10.1038/nbt.3597.

(62) Haug, K., Salek, R. M., Conesa, P., Hastings, J., et al. MetaboLights–an Open-Access General-Purpose Repository for Metabolomics Studies and Associated Meta-Data.*Nucleic Acids Research* **2013**, *41*, D781–D786, DOI: 10.1093/nar/gks1004.

(63) Orchard, S., Ammari, M., Aranda, B., Breuza, L., et al. The MIntAct Project—IntAct as a Common Curation Platform for 11 Molecular Interaction Databases.*Nucleic Acids Research* **2014**, *42*, D358–D363, DOI: 10.1093/nar/gkt1115.

(64) Orchard, S., Kerrien, S., Abbani, S., Aranda, B., et al. Protein Interaction Data Curation: The International Molecular Exchange (IMEx) Consortium.*Nature Methods* **2012**, *9*, 345–350, DOI: 10.1038/nmeth.1931.

(65) Perfetto, L., Pastrello, C., del-Toro, N., Duesbury, M., et al. The IMEx Coronavirus Interactome: An Evolving Map of Coronaviridae–Host Molecular Interactions.*Database* **2020**, *2020*, DOI: 10.1093/database/baaa096.

(66) Wu, A., Peng, Y., Huang, B., Ding, X., et al. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host & Microbe* **2020**, *27*, 325–328, DOI: 10.1016/j.chom.2020.02.001.

(67) Srinivasan, S., Cui, H., Gao, Z., Liu, M., et al. Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins. *Viruses* **2020**, *12*, 360, DOI: 10.3390/v12040360.

(68) Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181*, 271–280.e8, DOI: 10.1016/j.cell.2020.02.052.

(69) Struwe, W., Emmott, E., Bailey, M., Sharon, M., et al. The COVID-19 MS Coalition—Accelerating Diagnostics, Prognostics, and Treatment. *The Lancet* **2020**, *395*, 1761–1762, DOI: 10.1016/S0140-6736(20)31211-3.

(70) Dunham, W. H., Mullin, M., Gingras, A.-C. Affinity-Purification Coupled to Mass Spectrometry: Basic Principles and Strategies. *PROTEOMICS* **2012**, *12*, 1576–1590, DOI: 10.1002/pmic.201100523.

(71) Samavarchi-Tehrani, P., Samson, R., Gingras, A.-C. Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. *Molecular & Cellular Proteomics* **2020**, *19*, 757–773, DOI: 10.1074/mcp.R120.001941.

(72) Mateus, A., Kurzawa, N., Becher, I., Sridharan, S., et al. Thermal Proteome Profiling for Interrogating Protein Interactions. *Molecular Systems Biology* **2020**, *16*, DOI: 10.15252/msb.20199232.

(73) Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., et al. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **2015**, *162*, 425–440, DOI: 10.1016/j.cell.2015.06.043.

(74) Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., et al. Architecture of the Human Interactome Defines Protein Communities and Disease Networks. *Nature* **2017**, *545*, 505–509, DOI: 10.1038/nature22366.

(75) Huttlin, E. L., Bruckner, R. J., Navarrete-Perea, J., Cannon, J. R., et al. Dual Proteome-Scale Networks Reveal Cell-Specific Remodeling of the Human Interactome. *bioRxiv* **2020**, DOI: 10.1101/2020.01.19.905109.

(76) Nesvizhskii, A. I. Computational and Informatics Strategies for Identification of Specific Protein Interaction Partners in Affinity Purification Mass Spectrometry Experiments. *PROTEOMICS* **2012**, *12*, 1639–1655, DOI: 10.1002/pmic.201100537.

(77) Jeong, K., Kim, S., Bandeira, N. False Discovery Rates in Spectral Identification. *BMC Bioinformatics* **2012**, *13*, S2, DOI: 10.1186/1471-2105-13-S16-S2.

(78) Gillen, J., Nita-Lazar, A. Experimental Analysis of Viral–Host Interactions. *Frontiers in Physiology* **2019**, *10*, DOI: 10.3389/fphys.2019.00425.

(79) Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J.-P., et al. The CRAPome: A Contaminant Repository for Affinity Purification–Mass Spectrometry Data.*Nature Methods* **2013**, *10*, 730–736, DOI: 10.1038/nmeth.2557.

(80) Sowa, M. E., Bennett, E. J., Gygi, S. P., Harper, J. W. Defining the Human Deubiquitinating Enzyme Interaction Landscape.*Cell* **2009**, *138*, 389–403, DOI: 10.1016/j.cell.2009.04.042.

(81) Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A., et al. SAINT: Probabilistic Scoring of Affinity Purification–Mass Spectrometry Data.*Nature Methods* **2010**, *8*, 70–73, DOI: 10.1038/nmeth.1541.

(82) Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., et al. SAINTexpress: Improvements and Additional Features in Significance Analysis of INTeractome Software.*Journal of Proteomics* **2014**, *100*, 37–43, DOI: 10.1016/j.jprot.2013.10.023.

(83) Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., et al. Global Landscape of HIV–Human Protein Complexes.*Nature* **2012**, *481*, 365–370, DOI: 10.1038/nature10719.

(84) Titeca, K., Meysman, P., Gevaert, K., Tavernier, J., et al. SFINX: Straightforward Filtering Index for Affinity Purification–Mass Spectrometry Data Analysis.*Journal of Proteome Research* **2016**, *15*, 332–338, DOI: 10.1021/acs.jproteome.5b00666.

(85) Thomas, K., Benjamin, R.-K., Fernando, P., Brian, G., et al. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*; IOS Press: 2016, pp 87–90.

(86) Hulstaert, N., Shofstahl, J., Sachsenberg, T., Walzer, M., et al. ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion.*Journal of Proteome Research* **2020**, *19*, 537–542, DOI: 10.1021/acs.jproteome.9b00328.

(87) Eng, J. K., Jahan, T. A., Hoopmann, M. R. Comet: An Open-Source MS/MS Sequence Database Search Tool.*PROTEOMICS* **2013**, *13*, 22–24, DOI: 10.1002/pmic.201200439.

(88) McIlwain, S., Tamura, K., Kertesz-Farkas, A., Grant, C. E., et al. Crux: Rapid Open Source Protein Tandem Mass Spectrometry Analysis.*Journal of Proteome Research* **2014**, *13*, 4488–4491, DOI: 10.1021/pr500741y.

(89) Breuza, L., Poux, S., Estreicher, A., Famiglietti, M. L., et al. The UniProtKB Guide to the Human Proteome.*Database* **2016**, *2016*, bav120, DOI: 10.1093/database/bav120.

(90) Goloborodko, A. A., Levitsky, L. I., Ivanov, M. V., Gorshkov, M. V. Pyteomics-a Python Framework for Exploratory Data Analysis and Rapid Software Prototyping in Proteomics.*Journal of The American Society for Mass Spectrometry* **2013**, *24*, 301–304, DOI: 10.1007/s13361-012-0516-6.

(91) Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., et al. G:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update).*Nucleic Acids Research* **2019**, *47*, W191–W198, DOI: 10.1093/nar/gkz369.

(92) Orchard, S. Molecular Interaction Databases.*PROTEOMICS* **2012**, *12*, 1656–1662, DOI: 10.1002/pmic.201100484.

(93) Licata, L., Briganti, L., Peluso, D., Perfetto, L., et al. MINT, the Molecular Interaction Database: 2012 Update.*Nucleic Acids Research* **2011**, *40*, D857–D861, DOI: 10.1093/nar/gkr930.

(94) Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., et al. The BioGRID Interaction Database: 2019 Update.*Nucleic Acids Research* **2019**, *47*, D529–D541, DOI: 10.1093/nar/gky1079.

(95) Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., et al. STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets.*Nucleic Acids Research* **2018**, *47*, D607–D613, DOI: 10.1093/nar/gky1131.

(96) Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F. S. L., et al. PSICQUIC and PSISCORE: Accessing and Scoring Molecular Interactions.*Nature Methods* **2011**, *8*, 528–529, DOI: 10.1038/nmeth.1637.

(97) Hagberg, A. A., Schult, D. A., Swart, P. J. In *Proceedings of the 7th Python in Science Conference - SciPy '08*, ed. by Varoquaux, G., Vaught, T., Millman, J., Pasadena, CA USA, 2008, pp 11–15.

(98) Guy, R. K., DiPaola, R. S., Romanelli, F., Dutch, R. E. Rapid Repurposing of Drugs for COVID-19.*Science* **2020**, *368*, 829–830, DOI: 10.1126/science.abb9332.

(99) Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., et al. Discovery of SARS-CoV-2 Antiviral Drugs through Large-Scale Compound Repurposing.*Nature* **2020**, *586*, 113–119, DOI: 10.1038/s41586-020-2577-1.

(100) The RECOVERY Collaborative Group Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report.*New England Journal of Medicine* **2020**, DOI: 10.1056/NEJMoa2021436.

(101) Krammer, F. SARS-CoV-2 Vaccines in Development.*Nature* **2020**, DOI: 10.1038/s41586-020-2798-3.

(102) Du, L., He, Y., Zhou, Y., Liu, S., et al. The Spike Protein of SARS-CoV — a Target for Vaccine and Therapeutic Development.*Nature Reviews Microbiology* **2009**, *7*, 226–236, DOI: 10.1038/nrmicro2090.

(103) De Chassey, B., Meyniel-Schicklin, L., Vonderscher, J., André, P., et al. Virus-Host Interactomics: New Insights and Opportunities for Antiviral Drug Discovery.*Genome Medicine* **2014**, *6*, DOI: 10.1186/s13073-014-0115-1.

(104) Scott, D. E., Bayly, A. R., Abell, C., Skidmore, J. Small Molecules, Big Targets: Drug Discovery Faces the Protein–Protein Interaction Challenge.*Nature Reviews Drug Discovery* **2016**, *15*, 533–550, DOI: 10.1038/nrd.2016.29.

(105) Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., et al. DrugBank 5.0: A Major Update to the DrugBank Database for 2018.*Nucleic Acids Research* **2018**, *46*, D1074–D1082, DOI: 10.1093/nar/gkx1037.

(106) Wang, Y., Zhang, S., Li, F., Zhou, Y., et al. Therapeutic Target Database 2020: Enriched Resource for Facilitating Research and Early Development of Targeted Therapeutics.*Nucleic Acids Research* **2019**, DOI: 10.1093/nar/gkz981.

(107) Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., et al. BindingDB in 2015: A Public Database for Medicinal Chemistry, Computational Chemistry and Systems Pharmacology.*Nucleic Acids Research* **2016**, *44*, D1045–D1053, DOI: 10.1093/nar/gkv1072.

(108) Davies, M., Nowotka, M., Papadatos, G., Dedman, N., et al. ChEMBL Web Services: Streamlining Access to Drug Discovery Data and Utilities.*Nucleic Acids Research* **2015**, *43*, W612–W620, DOI: 10.1093/nar/gkv352.

(109) Deftereos, S. G., Giannopoulos, G., Vrachatis, D. A., Siasos, G. D., et al. Effect of Colchicine vs Standard Care on Cardiac and Inflammatory Biomarkers and Clinical Outcomes in Patients Hospitalized with Coronavirus Disease 2019: The GRECCO-19 Randomized clinicaltTrial.*JAMA Network Open* **2020**, *3*, e2013136, DOI: 10.1001/jamanetworkopen.2020.13136.

(110) Uddin, M. H., Zonder, J. A., Azmi, A. S. Exportin 1 Inhibition as Antiviral Therapy.*Drug Discovery Today* **2020**, *25*, 1775–1781, DOI: 10.1016/j.drudis.2020.06.014.

(111) Barros, R. O., Junior, F. L. C. C., Pereira, W. S., Oliveira, N. M. N., et al. Interaction of Drug Candidates with Various SARS-CoV-2 Receptors: An in Silico Study to Combat COVID-19.*Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00327.

(112) Richardson, P., Griffin, I., Tucker, C., Smith, D., et al. Baricitinib as Potential Treatment for 2019-nCoV Acute Respiratory Disease.*The Lancet* **2020**, *395*, e30–e31, DOI: 10.1016/S0140-6736(20)30304-4.

(113) Zeng, X., Song, X., Ma, T., Pan, X., et al. Repurpose Open Data to Discover Therapeutics for COVID-19 Using Deep Learning.*Journal of Proteome Research* **2020**, DOI: 10.1021/acs.jproteome.0c00316.

(114) Segler, M. H. S., Preuss, M., Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI.*Nature* **2018**, *555*, 604–610, DOI: 10.1038/nature25978.

(115) Stebbing, J., Phelan, A., Griffin, I., Tucker, C., et al. COVID-19: Combining Antiviral and Anti-Inflammatory Treatments.*The Lancet Infectious Diseases* **2020**, *20*, 400–402, DOI: 10.1016/S1473-3099(20)30132-8.

(116) Stebbing, J., Sánchez Nievas, G., Falcone, M., Youhanna, S., et al. JAK Inhibition Reduces SARS-CoV-2 Liver Infectivity and Modulates Inflammatory Responses to Reduce Morbidity and Mortality.*Science Advances* **2021**, *7*, eabe4724, DOI: 10.1126/sciadv.abe4724.

(117) Favalli, E. G., Biggioggero, M., Maioli, G., Caporali, R. Baricitinib for COVID-19: A Suitable Treatment?*The Lancet Infectious Diseases* **2020**, *20*, 1012–1013, DOI: 10.1016/S1473-3099(20)30262-0.

(118) Richardson, P. J., Corbellino, M., Stebbing, J. Baricitinib for COVID-19: A Suitable Treatment? – Authors' Reply.*The Lancet Infectious Diseases* **2020**, *20*, 1013–1014, DOI: 10.1016/S1473-3099(20)30270-X.

(119) Dawson, A. R., Mehle, A. Flu's Cues: Exploiting Host Post-Translational Modifications to Direct the Influenza Virus Replication Cycle.*PLOS Pathogens* **2018**, *14*, ed. by Evans, M. J., e1007205, DOI: 10.1371/journal.ppat.1007205.

(120) Fung, T. S., Liu, D. X. Post-Translational Modifications of Coronavirus Proteins: Roles and Function. *Future Virology* **2018**, *13*, 405–430, DOI: 10.2217/fvl-2018-0008.

(121) Song, H. C., Seo, M.-Y., Stadler, K., Yoo, B. J., et al. Synthesis and Characterization of a Native, Oligomeric Form of Recombinant Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein. *Journal of Virology* **2004**, *78*, 10328–10335, DOI: 10.1128/JVI.78.19.10328-10335.2004.

(122) Wu, C.-H., Yeh, S.-H., Tsay, Y.-G., Shieh, Y.-H., et al. Glycogen Synthase Kinase-3 Regulates the Phosphorylation of Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein and Viral Replication. *Journal of Biological Chemistry* **2009**, *284*, 5229–5239, DOI: 10.1074/jbc.M805747200.

(123) Harcourt, B. H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., et al. Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-like Protease Activity. *Journal of Virology* **2004**, *78*, 13600–13612, DOI: 10.1128/JVI.78.24.13600-13612.2004.

(124) Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., et al. The Papain-like Protease of Severe Acute Respiratory Syndrome Coronavirus Has Deubiquitinating Activity. *Journal of Virology* **2005**, *79*, 15189–15198, DOI: 10.1128/JVI.79.24.15189-15198.2005.

(125) Mielech, A. M., Kilianski, A., Baez-Santos, Y. M., Mesecar, A. D., et al. MERS-CoV Papain-like Protease Has deISGylating and Deubiquitinating Activities. *Virology* **2014**, *450-451*, 64–70, DOI: 10.1016/j.virol.2013.11.040.

(126) Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281–292.e6, DOI: 10.1016/j.cell.2020.02.058.

(127) Shajahan, A., Archer-Hartmann, S., Supekar, N. T., Gleinich, A. S., et al. Comprehensive Characterization of N- and O- Glycosylation of SARS-CoV-2 Human Receptor Angiotensin Converting Enzyme 2. *Glycobiology* **2020**, DOI: 10.1093/glycob/cwaa101.

(128) Shajahan, A., Supekar, N. T., Gleinich, A. S., Azadi, P. Deducing the N- and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology* **2020**, DOI: 10.1093/glycob/cwaa042.

(129) Sun, Z., Ren, K., Zhang, X., Chen, J., et al. Mass Spectrometry Analysis of Newly Emerging Coronavirus HCoV-19 Spike S Protein and Human ACE2 Reveals Camouflaging Glycans and Unique Post-Translational Modifications. *Engineering* **2020**, DOI: 10.1016/j.eng.2020.07.014.

(130) Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., et al. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* **2020**, *369*, 330–333, DOI: 10.1126/science.abb9983.

(131) Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., et al. Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe* **2020**, *28*, 586–601.e6, DOI: 10.1016/j.chom.2020.08.004.

(132) Supekar, N. T., Shajahan, A., Gleinich, A. S., Rouhani, D., et al. SARS-CoV-2 Nucleocapsid Protein Is Decorated with Multiple N- and O-Glycans. *bioRxiv* **2020**, DOI: 10.1101/2020.08.26.269043.

(133) Kim, M.-S., Zhong, J., Pandey, A. Common Errors in Mass Spectrometry-Based Analysis of Post-Translational Modifications.*PROTEOMICS* **2016**, *16*, 700–714, DOI: 10.1002/pmic.201500355.

(134) Bittremieux, W., Tabb, D. L., Impens, F., Staes, A., et al. Quality Control in Mass Spectrometry-Based Proteomics.*Mass Spectrometry Reviews* **2018**, *37*, 697–711, DOI: 10.1002/mas.21544.

(135) Tsur, D., Tanner, S., Zandi, E., Bafna, V., et al. Identification of Post-Translational Modifications by Blind Search of Mass Spectra.*Nature Biotechnology* **2005**, *23*, 1562–1567, DOI: 10.1038/nbt1168.

(136) Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., et al. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry–Based Proteomics.*Nature Methods* **2017**, *14*, 513–520, DOI: 10.1038/nmeth.4256.

(137) Chi, H., Liu, C., Yang, H., Zeng, W.-F., et al. Comprehensive Identification of Peptides in Tandem Mass Spectra Using an Efficient Open Search Engine.*Nature Biotechnology* **2018**, DOI: 10.1038/nbt.4236.

(138) Devabhaktuni, A., Lin, S., Zhang, L., Swaminathan, K., et al. TagGraph Reveals Vast Protein Modification Landscapes from Large Tandem Mass Spectrometry Datasets.*Nature Biotechnology* **2019**, *37*, 469–479, DOI: 10.1038/s41587-019-0067-5.

(139) Bittremieux, W., Meysman, P., Noble, W. S., Laukens, K. Fast Open Modification Spectral Library Searching through Approximate Nearest Neighbor Indexing.*Journal of Proteome Research* **2018**, *17*, 3463–3474, DOI: 10.1021/acs.jproteome.8b00359.

(140) Bittremieux, W., Laukens, K., Noble, W. S. Extremely Fast and Accurate Open Modification Spectral Library Searching of High-Resolution Mass Spectra Using Feature Hashing and Graphics Processing Units.*Journal of Proteome Research* **2019**, *18*, 3792–3799, DOI: 10.1021/acs.jproteome.9b00291.

(141) Solntsev, S. K., Shortreed, M. R., Frey, B. L., Smith, L. M. Enhanced Global Post-Translational Modification Discovery with MetaMorpheus.*Journal of Proteome Research* **2018**, *17*, 1844–1851, DOI: 10.1021/acs.jproteome.7b00873.

For table of contents only