Open modification searching of SARS-CoV-2–human protein interaction data reveals novel viral modification sites

Charlotte Adams[1,2], Kurt Boonen[2,3], Kris Laukens[1], Wout Bittremieux[4,*]

1. Department of Computer Science, University of Antwerp, 2020 Antwerpen, Belgium
2. Centre for Proteomics (CFP), University of Antwerp, 2020 Antwerpen, Belgium
3. Sustainable Health Department, Flemish Institute for Technological Research (VITO), 2400 Mol, Belgium
4. Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, United States of America

*Corresponding author: wbittremieux@health.ucsd.edu

Keywords

SARS-CoV-2; protein-protein interactions; open modification searching; post-translational modification; phosphorylation; ubiquitination; S-nitrosylation

Abstract

The outbreak of the SARS-CoV-2 coronavirus, the causative agent of the COVID-19 disease, has led to an ongoing global pandemic since 2019. Mass spectrometry can be used to understand the molecular mechanisms of viral infection by SARS-CoV-2, for example, by determining virus–host protein–protein interactions (PPIs) through which SARS-CoV-2 hijacks its human hosts during infection, and to study the role of post-translational modifications (PTMs). We have reanalyzed public affinity purification mass spectrometry data using open modification searching to investigate the presence of PTMs in the context of the SARS-CoV-2 virus–host PPI network. Based on an over two-fold increase in identified spectra, our detected protein interactions show a high overlap with independent mass spectrometry-based SARS-CoV-2 studies and virus–host interactions for alternative viruses, as well as previously unknown protein interactions. Additionally, we identified several novel modification sites on SARS-CoV-2 proteins that we investigated in relation to their interactions with host proteins. A detailed analysis of relevant modifications, including phosphorylation, ubiquitination, and S-nitrosylation, provides important hypotheses about the functional role of these modifications during viral infection by SARS-CoV-2.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly transmissible and pathogenic coronavirus that causes coronavirus disease 2019 (COVID-19). Since late 2019, it

has caused an ongoing global pandemic and socioeconomic crisis, with currently over 608 million people infected and over 6.5 million deaths globally (World Health Organization, September 2022) [1]. Despite intensive research by the scientific community, many important questions regarding its molecular mechanisms remain unanswered.

Viruses are opportunistic intracellular pathogens that depend on their interactions with host proteins to ensure their survival and propagation. The study of protein–protein interactions (PPIs) between SARS-CoV-2 and human proteins is important to understand the mechanisms of viral infection by SARS-CoV-2 [2] and develop therapeutic treatments [3]. Proteins and their function can be altered by post-translational modifications (PTMs) to increase the functional diversity of the limited number of proteins that are encoded. Interestingly, PTMs have been characterized on several coronavirus proteins, in spite of the fact that coronaviruses lack enzymes capable of introducing such modifications [4]. Elucidating the roles of PTMs in a mechanistic context is important to understand viral infection, as they are crucial for viral protein function and promote viral replication, assembly, and release. For example, phosphorylation of the SARS-CoV-2 nucleocapsid protein allows its shuttling between cellular compartments [5]. Additionally, insights in viral and host PTM dynamics offers a potential avenue towards the development of antiviral therapies. Removing PTMs that play a role in the enzymatic activity of viral proteins can aid the host in overcoming viral infection. Alternatively, viral proteins can be modified leading to their inactivation and/or proteasomal degradation, for example by attaching ubiquitin [6].

Affinity-purification mass spectrometry (AP-MS)-based proteomics was used to compile the first SARS-CoV-2–human PPI map, revealing interactions with proteins involved in major cellular processes, including DNA replication, RNA processing, and vesicle trafficking to give insights into SARS-CoV-2 infection [2]. Additionally, the interaction map revealed several human proteins that are targeted by existing drugs approved by the US Food and Drug Administration, which may be potential targets for drug repurposing.

The majority of previously conducted studies use standard sequence database searching to process AP-MS data [2, 7-10]. Most search engines require the user to explicitly specify potential protein modifications to be considered during searching. Unfortunately, taking into account multiple modifications simultaneously during spectrum identification is problematic. First, it leads to excessive search times due to the exponential increase in the number of candidate peptides that have to be considered. Second, it produces more random high-scoring matches due to the increase in candidates, leading to fewer identifications at a given false discovery rate (FDR) [11]. Therefore, a standard analysis typically only considers a handful of variable modifications. In the original analysis of the SARS-CoV-2 AP-MS data, only two modifications (N-terminal acetylation and methionine oxidation) were considered, leaving a substantial part of the data unexplored.

We have recently performed a computational reanalysis of the SARS-CoV-2 AP-MS data by Gordon et al. [2] to uncover additional virus–host interactions not reported in the original study, which highlighted further opportunities for drug repurposing [12]. However,

post-translational modifications were still not considered. Therefore, in this study we used an increasingly popular approach to overcome these limitations called "open modification searching" (OMS) to gain new insights into virus–host interactions. This innovative approach allows a modified spectrum to match against its unmodified variant by using a very wide precursor mass window. It is thus able to identify peptides carrying any type of modification. Algorithmic advances of the last few years now allow for the fast and accurate use of OMS, enabling an unbiased detection of modifications at an unparalleled scale [13-15]. The OMS solution ANN-SoLo [13, 14] was used to reprocess the SARS-CoV-2 AP-MS data by Gordon et al. [2]. The reanalysis resulted in a more than two-fold increase in identified spectra compared to the originally reported results, allowing more accurate PPI filtering. Additionally, several modified viral peptides were identified. Phosphorylation, ubiquitination, and S-nitrosylation were selected to be investigated in more detail, in view of their putative importance during a viral infection. For each of these PTMs, we detected novel PTM sites on SARS-CoV-2 proteins, revealing potential functional insights.

Experimental procedures

Experimental design and statistical rationale

A previously generated AP-MS dataset to study the SARS-CoV-2 virus–host interactome [2] was retrieved from the PRIDE repository (PXD018117) [16]. For the full experimental details, see the original study by Gordon et al. [2]. In brief, affinity purification was performed using 27 SARS-CoV-2 proteins that were individually tagged and expressed in triplicate (biological replicates) in HEK-293T cells. Bead-bound proteins were denatured, reduced, carbamidomethylated, and enzymatically digested using trypsin, and each sample was injected via an Easy-nLC 1200 (Thermo Fisher Scientific) into a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). The SARS-CoV-2 proteins that were included are: all mature nonstructural proteins (Nsps), except for Nsp3 and Nsp16; a mutated version of Nsp5 to disable its proteolytic activity (Nsp5_C145A); and all predicted SARS-CoV-2 open reading frames (Orfs), including the spike (S), membrane (M), nucleocapsid (N), and envelope (E) protein. Spectrum identifications were filtered at 1% FDR and PPIs were filtered using a SAINTexpress Bayesian false-discovery rate (BFDR) ≤ 0.05, an average spectral count ≥ 2, and a MiST score ≥ 0.7 (see further).

Spectrum identification using open modification searching

First, the downloaded raw files were converted to MGF files using ThermoRawFileParser (version 1.2.3) [17]. Next, OMS was performed using the ANN-SoLo spectral library search engine (version 0.2.4) [13, 14]. A combined human–SARS-CoV-2 spectral library was used for searching. The MassIVE-KB library (version 2018/06/15) was used as human spectral library. This is a comprehensive human HCD spectral library containing 2,154,269 unique precursors corresponding to 1,114,503 unique peptides, derived from publicly available mass spectrometry data in the MassIVE repository [18]. SARS-CoV-2 spectra were simulated by generating all possible tryptic peptide sequences from the SARS-CoV-2 protein sequences

downloaded from UniProt (version 2020/03/05) using Pyteomics (version 4.3.2) [19] and predicting the corresponding spectra using Prosit (version prosit_intensity_2020_hcd; collision energy 33 as determined by Prosit collision energy calibration) [20]. A simulated spectral library for the green fluorescent protein was generated in a similar fashion. A final spectral library was compiled by merging all spectra using SpectraST (version 5.0) [21] and adding decoy spectra in a 1:1 ratio using the shuffle-and-reposition method [22]. ANN-SoLo was configured to use a 20 ppm precursor mass tolerance during the first step of its cascade search and a 500 Da precursor mass tolerance during its open search. Other search settings were to filter peaks below 101 *m*/*z*, above 1500 *m*/*z*, and in a 0.5 *m*/*z* window around the precursor mass; a 0.02 *m*/*z* fragment mass tolerance; and a bin size of 0.05 *m*/*z*. The remaining settings were kept at their default values. Peptide-spectrum matches (PSMs) were filtered at 1% FDR using ANN-SoLo's built-in subgroup FDR procedure (Supplementary Table 1).

Additionally, OMS was performed using MSFragger (version 3.5) [15] and FragPipe (version 18.0) against a concatenated FASTA file containing human protein sequences (Uniprot reviewed sequences downloaded on 2020/02/28) [23], the SARS-CoV-2 protein sequences (version 2020/03/05), and the green fluorescent protein sequence. An equal number of decoy protein sequences were generated using FragPipe. The MSFragger search settings included a precursor mass tolerance between -150 Da and 500 Da, a fragment mass tolerance of 0.02 Da, and trypsin cleavage with up to two missed cleavages. Cysteine carbamidomethylation was used as a fixed modification, and oxidation of methionine and N-terminal acetylation were used as variable modifications. Other search settings were kept at their default values. PSMs were processed using PeptideProphet (version 4.4.0) [24] with the FragPipe default settings for open searches and filtered at 1% FDR.

Protein inference and protein-protein interaction filtering

Protein inference was performed using the Protein Inference Algorithms (PIA) tool (version 1.3.13) [25] based on Occam's razor. Only proteins with minimum two unique peptides were retained, while other settings were kept at their default values (Supplementary Table 2). Combined scoring of interacting proteins using SAINTexpress (version 3.6.3) [26] and MiST (https://modbase.compbio.ucsf.edu/mist/, version main.e2da2b0) [27] was used to filter high-confidence PPIs (Supplementary Table 3). Scoring thresholds were a SAINTexpress Bayesian false-discovery rate (BFDR) $\leq$ 0.05, an average spectral count $\geq$ 2, and a MiST score $\geq$ 0.7 (Supplementary Table 4).

To validate the filtered PPIs, a list of SARS-CoV-2–human interactions reported in seven alternative MS-based SARS-CoV-2 virus–host interactome studies [7-10, 28-30] were obtained from the BioGRID repository [31]. Additionally, PPIs from a previous reanalysis of the original AP-MS study were included [12]. Human interaction partners were queried in the VirHostNet database (version 3.0) [32] to investigate whether these proteins are also targeted by other viruses.

Gene ontology enrichment analysis

Gene Ontology (GO) enrichment analysis was performed for the human proteins that interact with each viral protein, using the enrichGO function of the clusterProfiler package (version 4.0.5) in R. Significant GO terms corresponding to the biological process category (1% Benjamini-Hochberg false discovery rate) were extracted and further refined to select non-redundant terms using the rrvgo package (version 1.4.4) with default parameters [33].

Investigation of post-translational modifications

The precursor mass differences observed in the ANN-SoLo results were referenced against the Unimod database [34] to determine the modifications that were present. PSMs that included specific PTMs (phosphorylation, ubiquitination, and S-nitrosylation) were manually investigated in more detail to disambiguate between alternative PTM assignments with near-identical mass and determine the modification site by visual inspection using the spectrum_utils Python package (version 0.3.3) [35].

Identified PTM sites were verified using PTM prediction tools and through literature study. Phosphorylation results were compared to three independent SARS-CoV-2 phosphoproteomic studies [36-38]. Phosphorylation site prediction was performed using NetPhos (version 3.1) [39, 40]. Additionally, GPS (version 5.0) [41] was used to predict kinase-specific phosphorylation sites (with a “high” threshold) for each of the SARS-CoV-2 proteins that were found to interact with known human kinases, retrieved from the KinHub database [42] (Supplementary Table 5–6). Ubiquitination results were compared to two independent SARS-CoV-2 ubiquitination studies [7, 43] and ubiquitination sites predicted by BDM-PUB (version 1.0) [44] (Supplementary Table 7). To our knowledge, to date there have been no S-nitrosylation sites reported on SARS-CoV-2. S-nitrosylation site prediction was performed with iSNO-PseAAC (version 1.0) [45], to validate observed S-nitrosylation sites (Supplementary Table 8–9).

Results

Increased spectrum identification rate using open modification searching boosts protein–protein interaction confidence

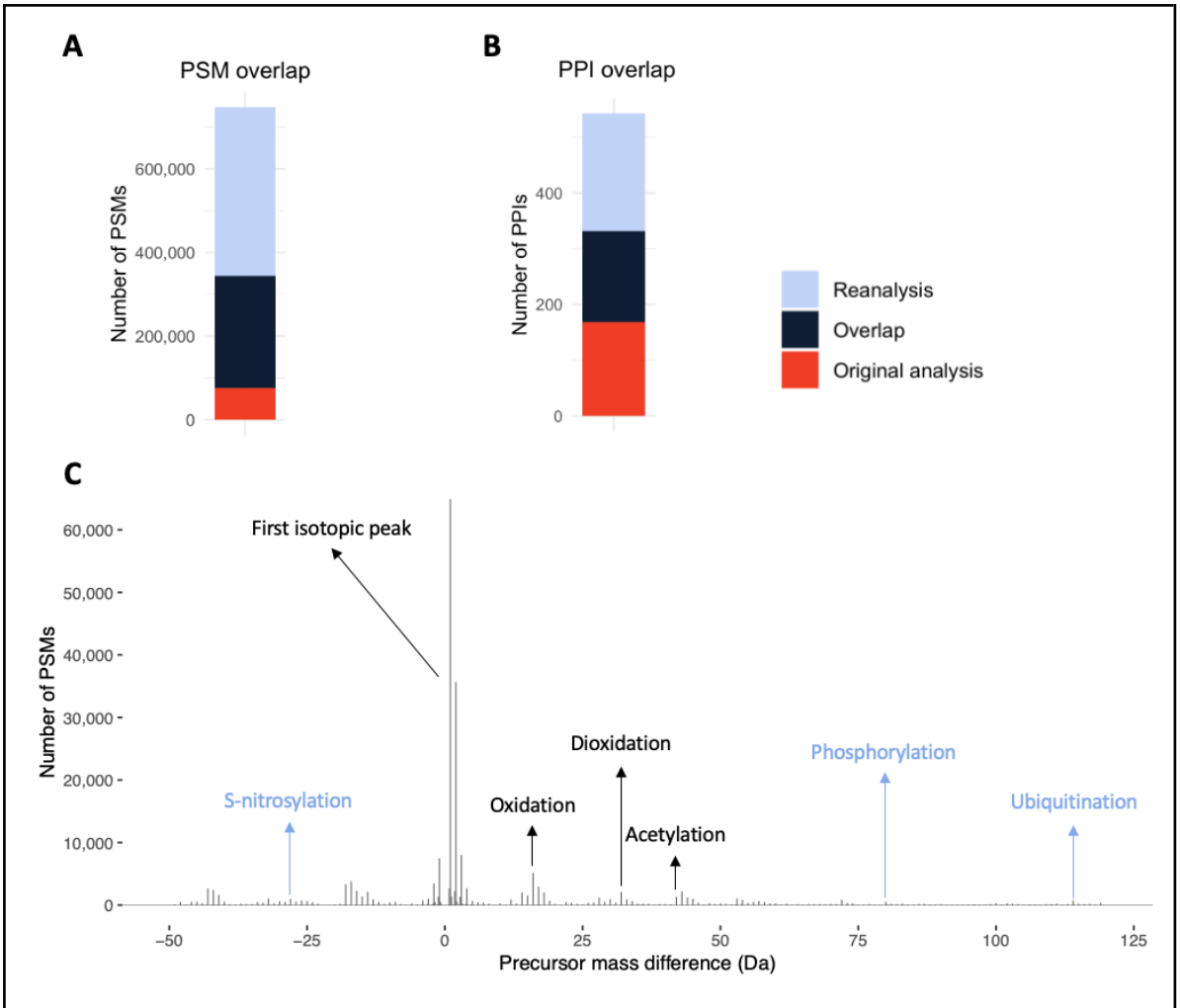

Figure 1. (A) Comparison in identification performance between the originally reported results using standard searching [2] and open modification searching. Open modification searching was able to identify more than twice as many spectra, corresponding to the identification of a large number of modified PSMs. **(B)** Comparison of the number of filtered PPIs. Half of the previously reported PPIs were reproduced [2], while 211 new PPIs were determined based on the extended identification results obtained using open modification searching. **(C)** Modifications can be derived from the precursor mass differences observed by open modification searching. Some of the most frequent delta masses are annotated with their likely modifications, sourced from Unimod [34], with modifications of artificial origin in black and relevant biological modifications in blue. The full list of observed precursor mass differences and their likely modifications is available in Supplementary Table 10.

Open modification searching using ANN-SoLo succeeded in identifying 830,743 of 2,503,010 total MS/MS spectra at 1% FDR (33% identification rate). This represents a 214% increase in identified spectra compared to the originally reported results [2] obtained by standard searching using MaxQuant [46] (Figure 1A). Notably, 402,586 PSMs correspond to modified peptides with non-zero precursor mass differences from open modification searching (Figure 1C, Supplementary Table 10). Besides the increase in identified spectra, an

important advantage of open modification searching is its ability to identify any type of PTM in an unbiased fashion, without the need to explicitly specify a limited number of variable modifications. This makes it possible to explore the general presence of PTMs in the context of the SARS-CoV-2 virus–host interactome. Frequently observed PTMs include modifications that were likely artificially introduced during sample processing [47], such as oxidation, dioxidation, and acetylation. Such ubiquitous modifications are typically included as variable modifications during standard searching. The OMS results also include unique observations of biologically relevant modifications at lower abundances, such as phosphorylation, ubiquitination, and S-nitrosylation.

Similar beneficial results can also be achieved with alternative OMS tools, such as MSFragger [15]. Performing an open search using MSFragger instead of ANN-SoLo succeeded in identifying 574,394 of 2,503,010 total MS/MS spectra at 1% FDR (23% identification rate). This represents a 148% increase in identified spectra compared to the originally reported results [2] (Supplementary Figure 1A). Furthermore, there is a strong correspondence in identification results between ANN-SoLo and MSFragger (Supplementary Figure 1B–C), which indicates the robustness of open modification searching, irrespective of the search engine that is employed. For simplicity of the downstream analyses and because ANN-SoLo was more sensitive than MSFragger, the ANN-SoLo results were used to investigate protein interactions during viral infection by SARS-CoV-2.

PPI filtering is crucial to separate true interactors from non-specific binders and contaminants. A combination of SAINTexpress and MiST filtering of the open modification searching results produced 375 high-confidence PPIs (Supplementary Table 4). These results contain 164 PPIs that overlap with the previously reported results [2] and 211 novel PPIs (Supplementary Figure 2). Notably, a previous reanalysis of these AP-MS data using alternative bioinformatics tools reported a similar overlap with the original PPI results [12]. The difference in detected PPIs is partly due to the difference in spectrum identification and PPI filtering strategies. In the original analysis a two-step filtering strategy was used. In the first step, the PPIs were filtered by a SAINTexpress BFDR ≤ 0.05, an average spectral count ≥ 2, and a MiST score ≥ 0.7. All proteins that fulfilled the first filtering step were searched in the CORUM [48] database of known protein complexes, and information was extracted about the stable protein complexes that they participated in. In the second step, all the proteins that formed known complexes with interactors identified in the first step were subjected to a filtering step with a lower stringency (MiST score ≥ 0.6) [2]. In contrast, in the reanalysis only a single filtering step was performed using a SAINTexpress BFDR ≤ 0.05, an average spectral count ≥ 2, and a MiST score ≥ 0.7.

When compared to the unfiltered PPI data of the original analysis [2], 199 of the 211 novel PPIs were previously detected as well but failed the original PPI filtering thresholds (Supplementary Figure 3). Most discarded PPIs (~88%) did not pass the minimum threshold for the MiST score, which is a linear combination of the prey abundance, the prey reproducibility across repeated runs, and the specificity of the prey relative to other baits [49]. About 28% of the discarded PPIs did not pass the SAINTexpress BFDR filter, which is based

on the abundance of the preys and control proteins [26]. The difference in filtered PPIs in the reanalysis is driven by the increased number of PSMs per identified protein from the open modification results, which resulted in a more reliable identification of both background proteins and potential interaction partners. This influences the PPI filtering and results in a more robust identification of true interactors.

The PPIs were validated against seven alternative MS-based SARS-CoV-2 virus–host interactome studies [7-10, 28-30] and the results from a previous reanalysis [12]. The largest overlap was found with the previous reanalysis and the original analysis, as these studies make use of the same experimental data and only differ in the computational tools that were employed (Figure 2A). Although most independent studies contributed unique PPIs, there is a strong correspondence in the superset of detected PPIs across all independent studies, with 317 of the 375 PPIs that were detected in this study previously reported in one or more of the other studies (Figure 2A, Supplementary Figure 2). Additionally, investigating the overlap with the VirHostNet database [32] shows that 328 of the human interaction partners are also targeted by other viruses (Figure 2B). These results are significantly enriched in the VirHostNet database (Fisher's exact test, p-value $\ll$ 0.001), which is expected as different viruses have similar strategies to hijack the host, for example, through functional conservation or viral recombination [32].

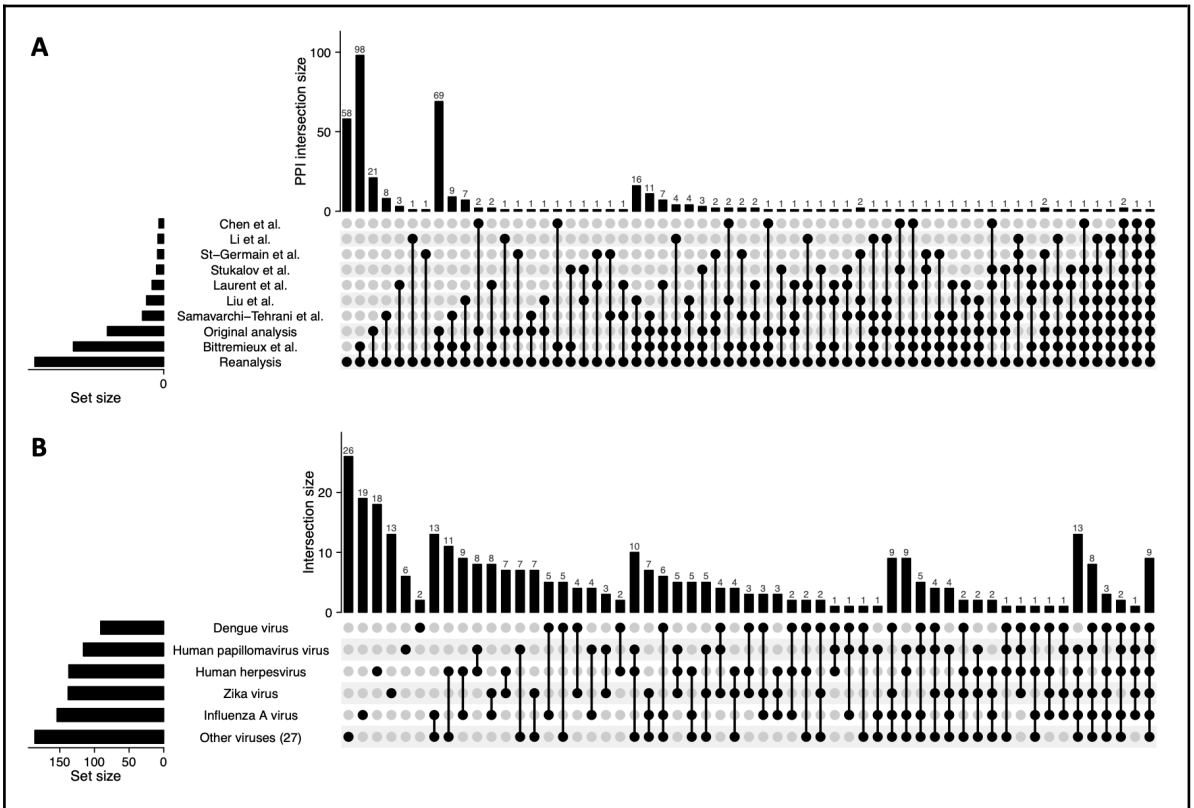

Figure 2. (A) Upset plot showing the overlap with seven alternative MS-based SARS-CoV-2 virus–host interactome studies, the original analysis, and a previous reanalysis of the AP-MS data performed by Bittremieux et al. [12]. We observe the largest overlap in PPIs with the previous reanalysis and the original analysis, as all are based on

the same AP-MS data. However, several novel detected PPIs have been observed in independent SARS-CoV-2 interactome studies as well, boosting the confidence in these interactions. **(B)** Upset plot showing the overlap of the detected human interaction partners with targets reported for different viruses in the VirHostNet database (excluding VirHostNet data for MERS-CoV, SARS-CoV-1, and SARS-CoV-2).

GO enrichment analysis based on the human interacting proteins for each SARS-CoV-2 protein indicates how viral infection might hijack major cellular processes, including metabolic processes involving non-coding RNA (ncRNA; Nsp8) and glycoproteins (Orf8), and RNA export from the nucleus (Orf6) (Figure 3). Both microRNA (miRNA) and long non-coding RNA (lncRNA) are ncRNA that can regulate gene expression and exhibit different expression profiles in COVID-19 patients compared to healthy people [50, 51]. Additionally, they have recently been found to act as viral modulators, regulating viral infection and host defense [52, 53]. Glycoprotein metabolic processes are expected to be relevant to Orf8, as induction of proinflammatory cytokine production by secreted Orf8 is glycosylation dependent [54]. RNA export from the nucleus is also relevant, as SARS-CoV-2-infected cells have an increased level of nuclear mRNA accumulation. Confirming our detected interactions between Orf6 and the mRNA export factors Rae1 and Nup98, a recent study found that Orf6 uses this process to trap host mRNA [55]. In addition to these major cellular processes, other biological processes relevant during viral infection were found, including viral transcription and viral gene expression (Supplementary Figure 4).

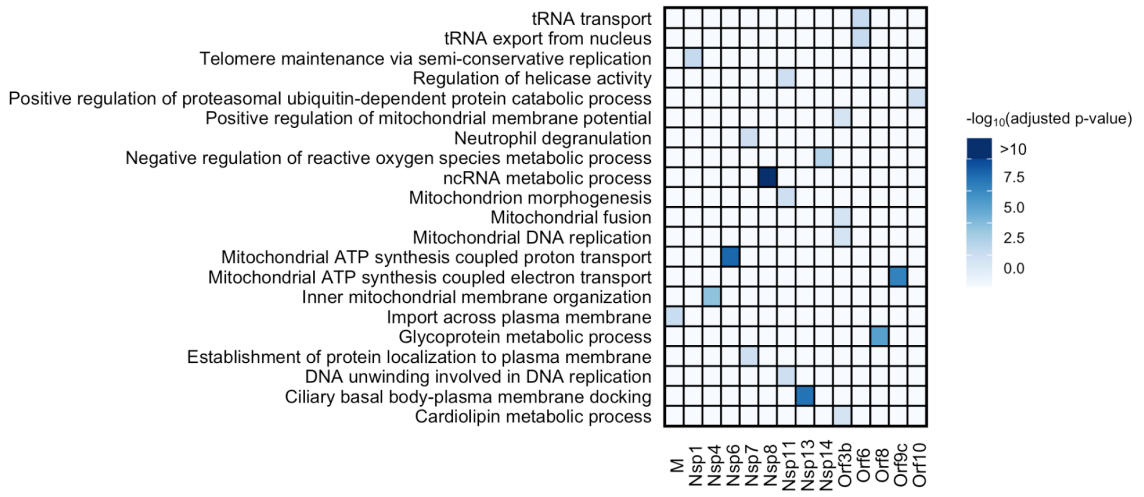

Figure 3. Heatmap showing the top GO terms from the GO enrichment analysis for each of the SARS-CoV-2 proteins.

Interaction between human kinases and phosphorylated viral proteins

Phosphorylation is a common PTM that impacts many basic cellular processes [56]. It usually results in a functional change of the target protein, interfering with its enzymatic activity,

cellular location, and/or association with other proteins. Several studies have indicated that also the function of viral proteins can be affected by their phosphorylation status. For example, association of the SARS-CoV-2 nucleocapsid protein with the 14-3-3 host proteins depends on its phosphorylation status [57] and phosphorylation of the SARS-CoV-1 nucleocapsid protein has been proven to be important for the regulation of the viral life cycle [58].

Twenty phosphorylated viral peptides were identified using open modification searching, and phosphosites were localized to specific residues using visual inspection where possible (Supplementary Table 5, Supplementary Figure 5–24). Notably, despite the fact that no phosphopeptide enrichment protocol was used during the original AP-MS study, we identified several phosphorylation sites that match results from specialized phosphoproteomics studies performed on SARS-CoV-2 infected cells [36-38] (Figure 4A).

Phosphorylation status is determined by the interplay between kinases and phosphatases. Kinases catalyze the attachment of phosphate groups to target proteins and phosphatases remove phosphate groups from target proteins [59]. SARS-CoV-2 does not encode any kinases or phosphatases, and thus relies on host proteins for its phosphorylation status. Several kinases and phosphatases have been identified as potential drug targets and a few kinase inhibitors are currently being used to treat COVID-19. For example, baricitinib is recommended by the World Health Organization (WHO) for patients with severe or critical COVID-19. It is a JAK inhibitor that suppresses overstimulation of the immune system by preventing phosphorylation of key proteins involved in the signal transduction that leads to immune activation and inflammation [60], and it is able to prevent SARS-CoV-2 from entering the cell by inhibiting clathrin-mediated endocytosis [61].

We detected multiple interactions between SARS-CoV-2 proteins and host kinases and phosphatases (Supplementary Tables 11–12), including RIPK1 (Nsp12) and TBK1 (Nsp13), both of which have previously been linked to SARS-CoV-2. Active phosphorylated RIPK1 was found in epithelial cell samples from COVID-19 patients [62], and a recent study discovered that Nsp12 promotes activation of RIPK1, which in turn enhances viral replication by stimulating the expression of viral receptors, such as ACE2 and EGFR [63]. Furthermore, Nsp13 is found to limit the activation of TBK1 by directly binding to TBK1 [64]. Additionally, we uncovered previously unknown interactions with host kinases, such as the interaction between Nsp12 and TLK1 (Figure 4B). This interaction is particularly interesting as kinase-specific phosphorylation site prediction revealed that an observed phosphorylation site on Nsp12 (T21) is most likely caused by TLK1.

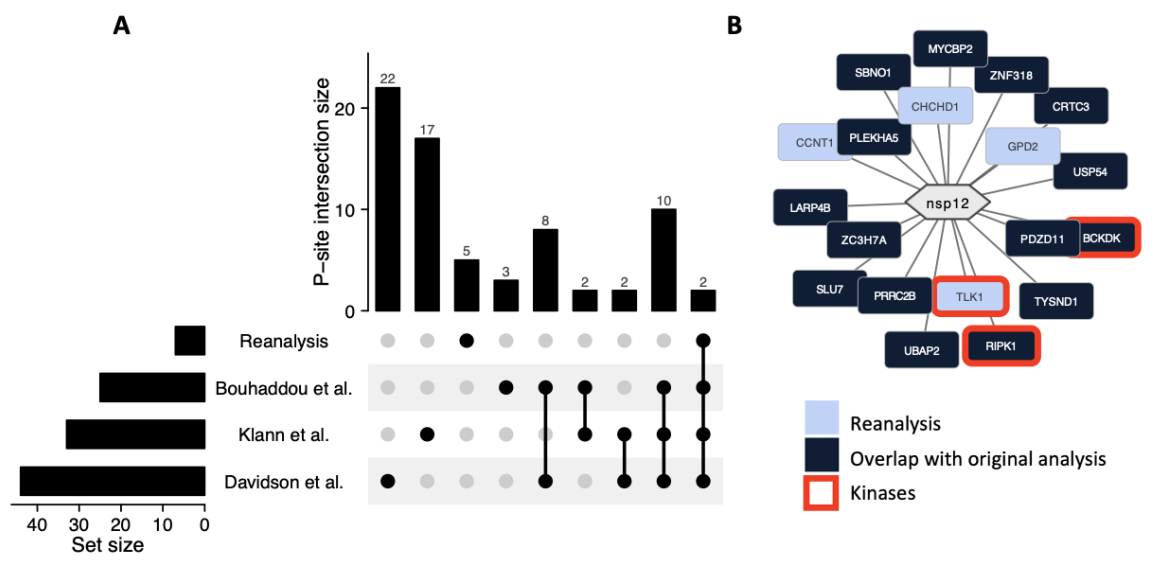

Figure 4. (A) Upset plot showing the overlap in phosphorylation sites between independent phosphoproteomic studies and the sites found in the current AP-MS reanalysis. **(B)** The PPI network of Nsp12 with highlighted kinases. Although RIPK1 and BCKDK were found in the original analysis, to our knowledge TLK1 has not been identified as an interaction partner of Nsp12 yet. After kinase-specific phosphorylation site prediction, we found that one of the observed phosphorylation sites (T21) is predicted to be phosphorylated by TLK1.

Ubiquitination of SARS-CoV-2 proteins

Ubiquitination is a common PTM that can affect the localization, stability, and function of proteins [65]. It regulates a variety of cellular processes, including protein degradation, protein trafficking, transcription, cell-cycle control, and cell signaling [66]. Depending on the cellular context, ubiquitin attachment may either promote or inhibit the viral life cycle. Viruses have developed means to exploit protein ubiquitination by enhancing or inhibiting ubiquitination of specific substrates depending on their needs. They can redirect protein degradation towards proteins with antiviral activity [67] or use ubiquitination to regulate viral proteins. For example, transcriptional function of the human immunodeficiency virus type-1 (HIV-1) Tat protein is increased by the addition of a single ubiquitin molecule [68]. More recently, it has been discovered that the inhibition of IFN-α signaling by SARS-CoV-2 Orf7a depends on the ubiquitination of K119 [69].

Ten ubiquitinated viral peptides were identified, and, where possible, ubiquitination sites were localized to specific residues using visual inspection (Supplementary Table 7, Supplementary Figure 25–34). Notably, all ubiquitination sites that could be confidently localized have been previously reported in independent studies [7, 43] (Figure 5A) or could be predicted. Interestingly, although ubiquitination of K16 on Orf9c has not been reported in SARS-CoV-2 yet, this site has been observed to be ubiquitinated in SARS-CoV-1 [7].

The final step of ubiquitination is performed by a ubiquitin (E3) ligase. This enzyme is particularly important as it determines to which substrate protein the ubiquitin is attached. A ubiquitin can be removed by deubiquitinating enzymes (DUBs). Both host E3 ligases and DUBs were found to interact with SARS-CoV-2 proteins (Supplementary Tables 13–14), suggesting that SARS-CoV-2 might hijack the host ubiquitination system. For example, Nsp13 was found to interact with the DUB USP13 (Figure 5B). According to a recent study, Nsp13 likely hijacks USP13 to prevent itself from degradation. Both knockdown of USP13 and its inhibition through spautin-1 resulted in decreased levels of Nsp13, suggesting that USP13 deubiquitinates and consequently stabilizes Nsp13 [70].

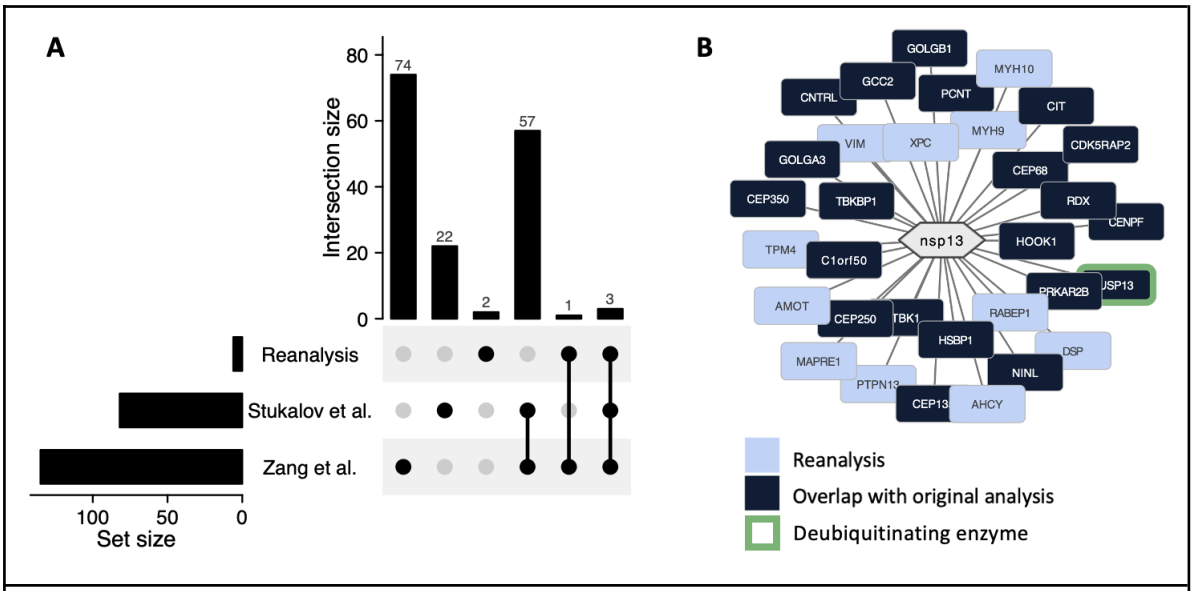

Figure 5. (A) Upset plot showing the overlap in ubiquitination sites between independent ubiquitination studies and the sites found in the current AP-MS reanalysis. The two novel ubiquitination sites were verified through *in silico* prediction. K16 on Orf9c has not been reported for SARS-CoV-2 yet, but was previously observed in SARS-CoV-1. **(B)** The PPI network of Nsp13 with the deubiquitinating enzyme USP13 highlighted.

S-nitrosylation of SARS-CoV-2 proteins

A promising compound currently undergoing clinical trials for COVID-19 is nitric oxide (NO) [71]. Besides its role as an important vasodilator, to prevent blood clot formation, it functions as a vital immune mediator, exerting broad spectrum antiviral effects [72]. Nitric oxide potentially prevents infection by SARS-CoV-2, as it was suggested for SARS-CoV [73]. The surprisingly low prevalence of smokers among hospitalized COVID-19 patients [74] could be attributed to the intermittent burst of high nitric oxide concentration in cigarette smoke [75]. Also nitrate-rich nutrition, exercise, and breathing through your nose is hypothesized to prevent a SARS-CoV-2 infection, as they all increase the nitric oxide concentration [73, 76].

The general antiviral mechanism appears to be the NO-mediated S-nitrosylation of viral and host proteins [77]. S-nitrosylation is the reversible, covalent attachment of nitric oxide to the thiol side chain of cysteine. It is one of the most important and universal PTMs, and it can act as a global regulator of protein function akin to phosphorylation and ubiquitination [78]. Interestingly, reactive cysteine residues are present in many viral and host proteins, representing possible targets for nitric oxide [77]. Results of an *in vitro* study suggest that S-nitrosylation of the SARS-CoV-2 3CL protease (Nsp5) can directly inhibit its protease activity and reduce viral replication [79].

To the best of our knowledge, there are currently no reported S-nitrosylation sites on SARS-CoV-2 proteins. Because cysteine residues are typically considered using a fixed carbamidomethylation modification (+57.021464 Da) introduced during sample processing by alkylation with iodoacetamide [47], Prosit [20]—which was used to simulate SARS-CoV-2 spectra—always considers cysteine residues to be carbamidomethylated. However, the presence of a prior modification can block reduction and alkylation of cysteine residues [80]. Therefore, rather than identifying S-nitrosylation directly based on its mass difference of +28.990164 Da, this modification is represented by a mass difference of -28.031300 Da (28.990164 Da - 57.021464) in the open modification searching results. Because the corresponding S-nitrosylation mass difference is identical to the mass difference observed from the valine to alanine and methionine to cysteine amino acid substitutions, careful visual inspection to localize the modification to specific amino acid residues was performed to confirm three S-nitrosylation sites (Supplementary Table 8, Supplementary Figure 35–52).

Interestingly, one of these S-nitrosylation sites (C45) is located on the main protease Nsp5, specifically, near the catalytic site. Nsp5 is one of two cysteine proteases necessary for viral replication and assembly, and its protease activity is suggested to be directly inhibited by S-nitrosylation [79]. A recent study found that C45 is a hyper-reactive cysteine with a higher nucleophilicity than C145—the catalytic cysteine of Nsp5—and identified C45 as an attractive binding site for the development of a covalent inhibitor [81].

Discussion

Mass spectrometry research can help to provide insights into the etiology of SARS-CoV-2 infection and identify potential therapeutic targets by investigating the roles of viral and host proteins during infection, their protein interactions, and post-translational modifications (PTMs) [12]. Additionally, it can be used to develop diagnostic strategies, such as through efforts of the CoV-MS consortium [82]—a partnership by multiple academic and industrial groups to increase applicability, accessibility, sensitivity, and robustness of a mass spectrometry-based diagnostic test that detects proteolytically digested SARS-CoV-2 proteins.

Here we have used open modification searching to reprocess SARS-CoV-2–human AP-MS data to investigate post-translational modifications in the context of protein–protein

interactions between SARS-CoV-2 and its human host. Studying these two aspects in tandem is especially relevant to understand the mechanisms of viral infection, because although PTMs are essential for viral replication, coronaviruses lack the enzymes to introduce these themselves. Whereas most SARS-CoV-2 studies so far have only considered a limited number of modifications commonly introduced during sample processing, our open modification searching strategy enabled the unbiased investigation of any PTM. This increased the number of identified spectra by more than two-fold, corresponding to newly identified modified peptides, and enabled us to put these PTMs in the context of the virus–host PPI network.

We discovered several combinations of relevant protein interactions and PTMs that hint at functional roles during viral infection. Specifically, we investigated phosphorylation, ubiquitination, and S-nitrosylation in more detail. These are reversible PTMs that are essential for a variety of cellular processes and could be of importance during viral infection. We found interactions between phosphorylated SARS-CoV-2 proteins and host kinases, between ubiquitinated SARS-CoV-2 proteins and host E3 ligases, and novel S-nitrosylation sites on SARS-CoV-2 proteins. Notably, even though no specialized modification enrichment was performed, we were able to confidently detect and localize multiple PTM sites that have been previously reported in independent enrichment studies, as well as novel PTM sites on SARS-CoV-2 proteins.

An important consideration in this study is the overlap in PPI results with the original study [2], which used standard searching instead of open modification searching. Although the vast majority of original spectrum identifications (79%) could be replicated using open modification searching, the overlap in PPIs was smaller (49%). Notably, another recent reanalysis of these data using alternative bioinformatics software tools showed a similarly limited overlap with the original PPI results [12]. Besides using stringent FDR control and other PPI filtering settings during all data processing steps, we validated the detected protein interactions through comparison with alternative MS-based SARS-CoV-2 studies and general virus–host interactions in the VirHostNet database. This showed a highly significant overlap between our PPI results and these independent sources, reinforcing the validity of the newly detected interactions.

The partial mismatch in PPI results can be largely explained by more complete identification of both background proteins and potential interaction partners using OMS. As such, the current study highlights the dependence of PPI filtering on the preceding protein identification results. First, open modification searching can be used to obtain high-quality identification data from which comprehensive PPI results can be obtained. Second, PPI filtering algorithms have to be able to robustly deal with uncertainty and missingness in the identification results when determining true protein interactions. Because OMS increased both the robustness of the PPI filtering results and enabled us to put PTM information in context of the protein interaction network, we suggest that this strategy is ideally suited for the analysis of MS-based PPI data.

In conclusion, we have used open modification searching to reanalyze open SARS-CoV-2 virus–host protein interaction data. The presented results enrich our knowledge of viral infection by SARS-CoV-2 by putting post-translational modifications in the context of the virus–host PPI network, which provides important hypotheses on the functional roles of PTMs during SARS-CoV-2 infection.

Data availability

All reanalysis results have been deposited to the ProteomeXchange Consortium [83] via the MassIVE repository with data set identifiers RMSV000000359.1 and RMSV000000309.4. Annotated spectra can be inspected in MS-Viewer (https://msviewer.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msviewer) using the key mcorgdznfy. All data processing code is freely available on GitHub as open source under the Apache 2.0 license at https://github.com/adamscharlotte/SARS-Cov-2-analysis.

Acknowledgements

This research was supported by the University of Antwerp BOF DOCPRO, FWO G070722N, and IBOF VLIR MIMICRY. KL acknowledges support of the University of Antwerp Core Facility BIOMINA. The computational resources and services used in this work were provided by the HPC core facility CalcUA of the University of Antwerp, and VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

References

1. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-19 in real time.* Lancet Infect Dis, 2020. **20**(5): p. 533-534.
2. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.
3. Valiente, P.A., et al., *Computational Design of Potent D-Peptide Inhibitors of SARS-CoV-2.* J Med Chem, 2021. **64**(20): p. 14955-14967.
4. Mishra, S., G. Bassi, and B.G. Nyomba, *Inter-proteomic posttranslational modifications of the SARS-CoV-2 and the host proteins A new frontier.* Exp Biol Med (Maywood), 2021. **246**(7): p. 749-757.
5. Tung, H.Y.L. and P. Limtung, *Mutations in the phosphorylation sites of SARS-CoV-2 encoded nucleocapsid protein and structure model of sequestration by protein 14-3-3.* Biochem Biophys Res Commun, 2020. **532**(1): p. 134-138.
6. Kumar, R., et al., *Role of Host-Mediated Post-Translational Modifications (PTMs) in RNA Virus Pathogenesis.* Int J Mol Sci, 2020. **22**(1).
7. Stukalov, A., et al., *Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV.* Nature, 2021. **594**(7862): p. 246-252.
8. Li, J., et al., *Virus-Host Interactome and Proteomic Survey Reveal Potential Virulence Factors Influencing SARS-CoV-2 Pathogenesis.* Med (New York, N.y.), 2020.

9. Chen, Z., et al., *Comprehensive analysis of the host-virus interactome of SARS-CoV-2.* bioRxiv, 2021: p. 2020.12.31.424961.
10. Liu, X., et al., *SARS-CoV-2-host proteome interactions for antiviral drug discovery.* Mol Syst Biol, 2021. **17**(11): p. e10396.
11. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nat Methods, 2007. **4**(3): p. 207-14.
12. Bittremieux, W., et al., *Open Science Resources for the Mass Spectrometry-Based Analysis of SARS-CoV-2.* J Proteome Res, 2021. **20**(3): p. 1464-1475.
13. Bittremieux, W., et al., *Fast Open Modification Spectral Library Searching through Approximate Nearest Neighbor Indexing.* J Proteome Res, 2018. **17**(10): p. 3463-3474.
14. Bittremieux, W., K. Laukens, and W.S. Noble, *Extremely Fast and Accurate Open Modification Spectral Library Searching of High-Resolution Mass Spectra Using Feature Hashing and Graphics Processing Units.* J Proteome Res, 2019. **18**(10): p. 3792-3799.
15. Kong, A.T., et al., *MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics.* Nat Methods, 2017. **14**(5): p. 513-520.
16. Perez-Riverol, Y., et al., *The PRIDE database and related tools and resources in 2019: improving support for quantification data.* Nucleic Acids Res, 2019. **47**(D1): p. D442-D450.
17. Hulstaert, N., et al., *ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion.* J Proteome Res, 2020. **19**(1): p. 537-542.
18. Wang, M., et al., *Assembling the Community-Scale Discoverable Human Proteome.* Cell Syst, 2018. **7**(4): p. 412-421 e5.
19. Levitsky, L.I., et al., *Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework.* J Proteome Res, 2019. **18**(2): p. 709-714.
20. Gessulat, S., et al., *Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning.* Nat Methods, 2019. **16**(6): p. 509-518.
21. Lam, H., et al., *Development and validation of a spectral library searching method for peptide identification from MS/MS.* Proteomics, 2007. **7**(5): p. 655-67.
22. Lam, H., E.W. Deutsch, and R. Aebersold, *Artificial decoy spectral libraries for false discovery rate estimation in spectral library searching in proteomics.* J Proteome Res, 2010. **9**(1): p. 605-10.
23. Breuza, L., et al., *The UniProtKB guide to the human proteome.* Database (Oxford), 2016. **2016**.
24. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.
25. Uszkoreit, J., et al., *PIA: An Intuitive Protein Inference Engine with a Web-Based User Interface.* J Proteome Res, 2015. **14**(7): p. 2988-97.
26. Teo, G., et al., *SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software.* J Proteomics, 2014. **100**: p. 37-43.
27. Jager, S., et al., *Global landscape of HIV-human protein complexes.* Nature, 2011. **481**(7381): p. 365-70.
28. Samavarchi-Tehrani, P., et al., *A SARS-CoV-2–host proximity interactome.* bioRxiv, 2020.

29. St-Germain, J., 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.
30. Laurent, E., 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.
31. Oughtred, R., et al., *The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions.* Protein Sci, 2021. **30**(1): p. 187-200.
32. Guirimand, T., S. Delmotte, and V. Navratil, *VirHostNet 2.0: surfing on the web of virus/host molecular interactions data.* Nucleic Acids Res, 2015. **43**(Database issue): p. D583-7.
33. Sayols, S., *rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms.* 2020.
34. Creasy, D.M. and J.S. Cottrell, *Unimod: Protein modifications for mass spectrometry.* Proteomics, 2004. **4**(6): p. 1534-6.
35. Bittremieux, W., *spectrum_utils: A Python Package for Mass Spectrometry Data Processing and Visualization.*Anal Chem, 2020. **92**(1): p. 659-661.
36. Klann, K., et al., *Growth Factor Receptor Signaling Inhibition Prevents SARS-CoV-2 Replication.* Mol Cell, 2020. **80**(1): p. 164-174 e4.
37. Davidson, A.D., et al., *Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein.* Genome Med, 2020. **12**(1): p. 68.
38. Bouhaddou, M., et al., *The Global Phosphorylation Landscape of SARS-CoV-2 Infection.* Cell, 2020. **182**(3): p. 685-712 e19.
39. Blom, N., S. Gammeltoft, and S. Brunak, *Sequence and structure-based prediction of eukaryotic protein phosphorylation sites.* J Mol Biol, 1999. **294**(5): p. 1351-62.
40. Blom, N., et al., *Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence.* Proteomics, 2004. **4**(6): p. 1633-49.
41. Wang, C., et al., *GPS 5.0: An Update on the Prediction of Kinase-specific Phosphorylation Sites in Proteins.*Genomics Proteomics Bioinformatics, 2020. **18**(1): p. 72-80.
42. Eid, S., et al., *KinMap: a web-based tool for interactive navigation through human kinome data.* BMC Bioinformatics, 2017. **18**(1): p. 16.
43. Zhang, H., et al., *Ubiquitin-Modified Proteome of SARS-CoV-2-Infected Host Cells Reveals Insights into Virus-Host Interaction and Pathogenesis.* J Proteome Res, 2021. **20**(5): p. 2224-2239.
44. Li X., G.X., Ren J., Jin C., Xue Y., *BDM-PUB: Computational prediction of protein ubiquitination sites with a Bayesian discriminant method.* 2009.
45. Xu, Y., et al., *iSNO-PseAAC: predict cysteine S-nitrosylation sites in proteins by incorporating position specific amino acid propensity into pseudo amino acid composition.* PLoS One, 2013. **8**(2): p. e55844.
46. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.

47. Bittremieux, W., et al., *Quality control in mass spectrometry-based proteomics.* Mass Spectrom Rev, 2018. **37**(5): p. 697-711.
48. Giurgiu, M., et al., *CORUM: the comprehensive resource of mammalian protein complexes-2019.* Nucleic Acids Res, 2019. **47**(D1): p. D559-D563.
49. Verschueren, E., et al., *Scoring Large-Scale Affinity Purification Mass Spectrometry Datasets with MiST.* Curr Protoc Bioinformatics, 2015. **49**: p. 8 19 1-8 19 16.
50. Li, C., et al., *Differential microRNA expression in the peripheral blood from human patients with COVID-19.* J Clin Lab Anal, 2020. **34**(10): p. e23590.
51. Wu, Y., et al., *A study of differential circRNA and lncRNA expressions in COVID-19-infected peripheral blood.* Sci Rep, 2021. **11**(1): p. 7991.
52. Li, W., et al., *Epstein barr virus encodes miRNAs to assist host immune escape.* J Cancer, 2020. **11**(8): p. 2091-2100.
53. Liao, Y., et al., *Host Non-Coding RNA Regulates Influenza A Virus Replication.* Viruses, 2022. **14**(1): p. 51.
54. Wu, X., et al., *Secreted ORF8 is a pathogenic cause of severe Covid-19 and potentially targetable with select NLRP3 inhibitors.* bioRxiv, 2021: p. 2021.12.02.470978.
55. Addetia, A., et al., *SARS-CoV-2 ORF6 Disrupts Bidirectional Nucleocytoplasmic Transport through Interactions with Rae1 and Nup98.* mBio, 2021. **12**(2).
56. Ubersax, J.A. and J.E. Ferrell, Jr., *Mechanisms of specificity in protein phosphorylation.* Nat Rev Mol Cell Biol, 2007. **8**(7): p. 530-41.
57. Tugaeva, K.V., et al., *The Mechanism of SARS-CoV-2 Nucleocapsid Protein Recognition by the Human 14-3-3 Proteins.* J Mol Biol, 2021. **433**(8): p. 166875.
58. Wu, C.H., et al., *Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication.* J Biol Chem, 2009. **284**(8): p. 5229-39.
59. Bononi, A., et al., *Protein kinases and phosphatases in the control of cell fate.* Enzyme Res, 2011. **2011**: p. 329098.
60. McInnes, I.B., et al., *Comparison of baricitinib, upadacitinib, and tofacitinib mediated regulation of cytokine signaling in human leukocyte subpopulations.* Arthritis Res Ther, 2019. **21**(1): p. 183.
61. Stebbing, J., et al., *COVID-19: combining antiviral and anti-inflammatory treatments.* Lancet Infect Dis, 2020. **20**(4): p. 400-402.
62. Riebeling, T., et al., *Primidone blocks RIPK1-driven cell death and inflammation.* Cell Death Differ, 2021. **28**(5): p. 1610-1626.
63. Xu, G., et al., *SARS-CoV-2 promotes RIPK1 activation to facilitate viral propagation.* Cell Res, 2021. **31**(12): p. 1230-1243.
64. Chathuranga, K., et al., *Regulation of antiviral innate immune signaling and viral evasion following viral genome sensing.* Exp Mol Med, 2021. **53**(11): p. 1647-1668.
65. Hershko, A. and A. Ciechanover, *The ubiquitin system.* Annu Rev Biochem, 1998. **67**: p. 425-79.
66. Pickart, C.M., *Back to the future with ubiquitin.* Cell, 2004. **116**(2): p. 181-90.
67. Lindner, H.A., *Deubiquitination in virus infection.* Virology, 2007. **362**(2): p. 245-56.
68. Bres, V., et al., *A non-proteolytic role for ubiquitin in Tat-mediated transactivation of the HIV-1 promoter.* Nat Cell Biol, 2003. **5**(8): p. 754-61.

69. Cao, Z., et al., *Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response.* Cell Mol Immunol, 2021. **18**(3): p. 746-748.
70. Guo, G., et al., *SARS-CoV-2 non-structural protein 13 (nsp13) hijacks host deubiquitinase USP13 and counteracts host antiviral immune response.* Signal Transduct Target Ther, 2021. **6**(1): p. 119.
71. Bhatia, V., L. Elnagary, and S. Dakshinamurti, *Tracing the path of inhaled nitric oxide: Biological consequences of protein nitrosylation.* Pediatr Pulmonol, 2021. **56**(2): p. 525-538.
72. Akaike, T., *Role of free radicals in viral pathogenesis and mutation.* Rev Med Virol, 2001. **11**(2): p. 87-101.
73. Ignarro, L.J., *Inhaled NO and COVID-19.* Br J Pharmacol, 2020. **177**(16): p. 3848-3849.
74. Meini, S., et al., *The Paradox of the Low Prevalence of Current Smokers Among COVID-19 Patients Hospitalized in Nonintensive Care Wards: Results From an Italian Multicenter Case-Control Study.* Nicotine Tob Res, 2021. **23**(8): p. 1436-1440.
75. Hedenstierna, G., et al., *Nitric oxide dosed in short bursts at high concentrations may protect against Covid 19.*Nitric Oxide, 2020. **103**: p. 1-3.
76. Swathi Krishna, S., A. Thennavan, and S.K. Kanthlal, *Dietary foods containing nitric oxide donors can be early curators of SARS-CoV-2 infection: A possible role in the immune system.* J Food Biochem, 2021: p. e13884.
77. Colasanti, M., et al., *S-nitrosylation of viral proteins: molecular bases for antiviral effect of nitric oxide.* IUBMB Life, 1999. **48**(1): p. 25-31.
78. Mnatsakanyan, R., et al., *Proteome-wide detection of S-nitrosylation targets and motifs using bioorthogonal cleavable-linker-based enrichment and switch technique.* Nat Commun, 2019. **10**(1): p. 2195.
79. Akaberi, D., et al., *Mitigation of the replication of SARS-CoV-2 by nitric oxide in vitro.* Redox Biol, 2020. **37**: p. 101734.
80. Liu, M., et al., *Site-specific proteomics approach for study protein S-nitrosylation.* Anal Chem, 2010. **82**(17): p. 7160-8.
81. Verma, N., J.A. Henderson, and J. Shen, *Proton-Coupled Conformational Activation of SARS Coronavirus Main Proteases and Opportunity for Designing Small-Molecule Broad-Spectrum Targeted Covalent Inhibitors.* J Am Chem Soc, 2020. **142**(52): p. 21883-21890.
82. Van Puyvelde, B., et al., *Cov-MS: A Community-Based Template Assay for Mass-Spectrometry-Based Protein Detection in SARS-CoV-2 Patients.* JACS Au, 2021. **1**(6): p. 750-765.
83. Deutsch, E.W., et al., *The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition.* Nucleic Acids Res, 2017. **45**(D1): p. D1100-D1106.