Supplemental methods

Supplemental Figure 1 Method: All MS runs were compared and clustered using standard artMS (https://github.com/biodavidjm/artMS) procedures on observed feature intensities computed by MaxQuant. Supplemental Figure 1 shows all Pearson’s pairwise correlations between MS runs, and are clustered according to similar correlation patterns.

Supplemental Figure 2 Method: See main text.

Supplemental Figure 3 Method: PFAM domain enrichment analysis. The enrichment of individual PFAM domains (or PFAM clans) \(^1\) was calculated with a hypergeometric test where success is defined as number of domains, and the number of trials is the number of individual preys pulled-down with each viral bait. The population values were the numbers of individual PFAM domains and clans in the human proteome. To make sure that the p-values that signify enrichment were meaningful, we only considered PFAM domains that have been pulled-down at least three times with any SARS-CoV-2 protein, and which occur in the human proteome at least five times. In SI Figure 3 we show PFAM domains/clans with the lowest p-value for a given viral bait protein.

Supplemental Figure 4 and 5 Method: Expression analysis of interacting genes. We used GTEx (version 8, median gene-level transcripts per million (TPM) by tissue), which consisted of 17382 samples (578 lung samples) \(^2\) to examine the mRNA expression of all interacting proteins (n=323). The comparison gene group was all RefSeq genes (n=24,491). The lung expression values represent the median expression of each gene across the GTEx lung samples. The lung enrichment values are calculated by dividing the median expression of each gene in lung tissue by the median expression of each gene across all tissues (including lung). A value of greater than one indicates that the gene expression is enriched in lung tissue. Values were plotted on a log10 scale. All figures and statistics were produced in Python3 and code and reference tables can be found at: (https://github.com/stephaniewanko/Fraser_Lab/tree/master/QCRG_COVID19).

Supplemental Figure 6 Method: Conservation analysis of interacting genes. We used gnomAD version 2.1 \(^3\), which consists of 125,748 exomes and 15,708 genomes, to determine human genetic variation observed in the interacting proteins (n=323) versus all Refseq genes (n=24,491). Briefly the observed/expected ratio per gene indicates the number of observed variants of that type divided by the number of expected mutations of that type, with a lower observed/expected ratio indicating strong intolerance toward mutation. The number of expected variants were estimated based on the number of CpG and non-CpG transitions observed across the genome \(^4\). All figures and statistics were produced in Python3 and code and reference tables can be found at: (https://github.com/stephaniewanko/Fraser_Lab/tree/master/QCRG_COVID19).

Supplemental Figure 7 Method: Nsp5 main protease (3Clpro) cleavage prediction. We used sequence specificity data for SARS nsp5 \(^5\) (98.7% identical to SARS-CoV-2 nsp5) and NetCorona \(^5\) to predict cleavage sites within interacting factors. PDB ID: 1UJ1 served as template for peptide docking which was performed using the predicted P4-P1 residues (BioLuminate, Schrödinger, LLC). Illustration of the docked model was generated in PyMol (Schrödinger, LLC).
Supplemental Figure 8 Method: Orf6 consensus sequence analysis. Orf6 sequence homologs were identified using the BLAST tool (accession number YP_009724394.1), run with the default settings: gap opening and extension costs of 11 and 1, respectively, BLOSUM62 as the scoring matrix, and an e-value threshold of 10. The search yielded 34 homologous sequences. The multiple sequence alignment was visualized using the MView web server: https://www.ebi.ac.uk/Tools/msa/mview/ and the WebLogo server (https://weblogo.berkeley.edu/logo.cgi).

Supplemental Table 1 Method: See main text.

Supplemental Table 2 Method: See main text.

Supplemental Figure 3-6 Method: Chemoinformatic Analysis of SARS-CoV2 Interacting Partners. To identify drugs and reagents that modulate the 332 host factors interacting with SARS-CoV-2-HEK293T (MIST >= 0.70), we used two approaches: 1) a chemoinformatic analysis of open-source chemical databases and 2) a target- and pathway-specific literature search, drawing on specialist knowledge within our group. Chemoinformatically, we retrieved 2,472 molecules from the IUPHAR/BPS Guide to Pharmacology (2020-3-12) that interacted with 30 human "prey" proteins (38 approved, 71 in clinical trials), and found 10,883 molecules (95 approved, 369 in clinical trials) from the ChEMBL25 database (Supplementary Tables 5, 6). For both approaches, molecules were prioritized on their FDA approval status, activity at the target of interest better than 1 μM, and commercial availability, drawing on the ZINC database. FDA approved molecules were prioritized except when clinical candidates or preclinical research molecules had substantially better selectivity or potency on-target. In some cases, we considered molecules with indirect mechanisms of action on the general pathway of interest based solely on literature evidence (e.g., captopril modulates ACE2 indirectly via its direct interaction with Angiotensin Converting Enzyme, ACE). Finally, we predicted 6 additional molecules (2 approved, 1 in clinical trials) for proteins with MIST scores between 0.7-0.6 to viral baits (Supplemental Tables 3 and 4). Complete methods can be found here (www.github.com/momeara/BioChemPantry/vignette/COVID19).

Supplementary Methods References

