Article

A proteome-scale map of the SARS-CoV-2– human contactome

A full list of authors and their affiliations appears at the end of the paper.

Understanding the mechanisms of coronavirus disease 2019 (COVID-19) disease severity to efficiently design therapies for emerging virus variants remains an urgent challenge of the ongoing pandemic. Infection and immune reactions are mediated by direct contacts between viral molecules and the host proteome, and the vast majority of these virus–host contacts (the 'contactome') have not been identified. Here, we present a systematic contactome map of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with the human host encompassing more than 200 binary virus–host and intraviral protein–protein interactions. We find that host proteins genetically associated with comorbidities of severe illness and long COVID are enriched in SARS-CoV-2 targeted network communities. Evaluating contactome-derived hypotheses, we demonstrate that viral NSP14 activates nuclear factor κB (NF-κB)-dependent transcription, even in the presence of cytokine signaling. Moreover, for several tested host proteins, genetic knock-down substantially reduces viral replication. Additionally, we show for USP25 that this effect is phenocopied by the small-molecule inhibitor AZ1. Our results connect viral proteins to human genetic architecture for COVID-19 severity and offer potential therapeutic targets.

Despite over 200,000 SARS-CoV-2 publications in the past two years, fundamental questions remain about the molecular mechanisms of genetic risk factors for severe and fatal COVID-19, the cause of long-persisting disease symptoms (long COVID) and the challenge to identify therapeutic targets[1]. These issues remain urgent in light of incomplete vaccination rates, continuously emerging variants and anticipated future pathogens. Fundamentally, infections are initiated by physical contacts between viral proteins and cellular receptors that set off molecular rearrangements culminating in viral entry and unpacking, followed by cellular reprogramming and host defense response triggering. Each of these steps is mediated by contacts between viral and host molecules that determine functional consequences, including proteolytic cleavage or inflammatory signaling, and ultimately clinical manifestations (Fig. 1a). Therefore, understanding the mechanisms by which human genetic variation affects COVID-19, as well as the behavior of newly emerging virus variants such as Delta (δ) and Omicron (o),

requires knowledge of these contacts to enable studies on how variants functionally alter virus–host interactions. For SARS-CoV-2, the contacts between the viral spike and human ACE2 proteins are documented by several hundred structures. In contrast, no direct interaction partners are known for many other viral proteins, precluding even domain-level contact models. Because co-complex assays predominantly detect indirect protein-associations[2], the virus–host contactome remains largely unexplored and unknown. To address this fundamental research gap, we systematically identified protein–protein contacts between SARS-CoV-2 and the human proteome.

Results

SARS-CoV-2–host contactome mapping
We used a multiassay screening and evaluation framework to generate a high-quality virus–host contactome map[2,3]. To increase detection sensitivity in the initial screening by yeast two hybrid (Y2H), we

✉e-mail: caroline.demeret@pasteur.fr; marc_vidal@dfci.harvard.edu; michael_calderwood@dfci.harvard.edu; fritz.roth@utoronto.ca; pascal.falter-braun@helmholtz-muenchen.de
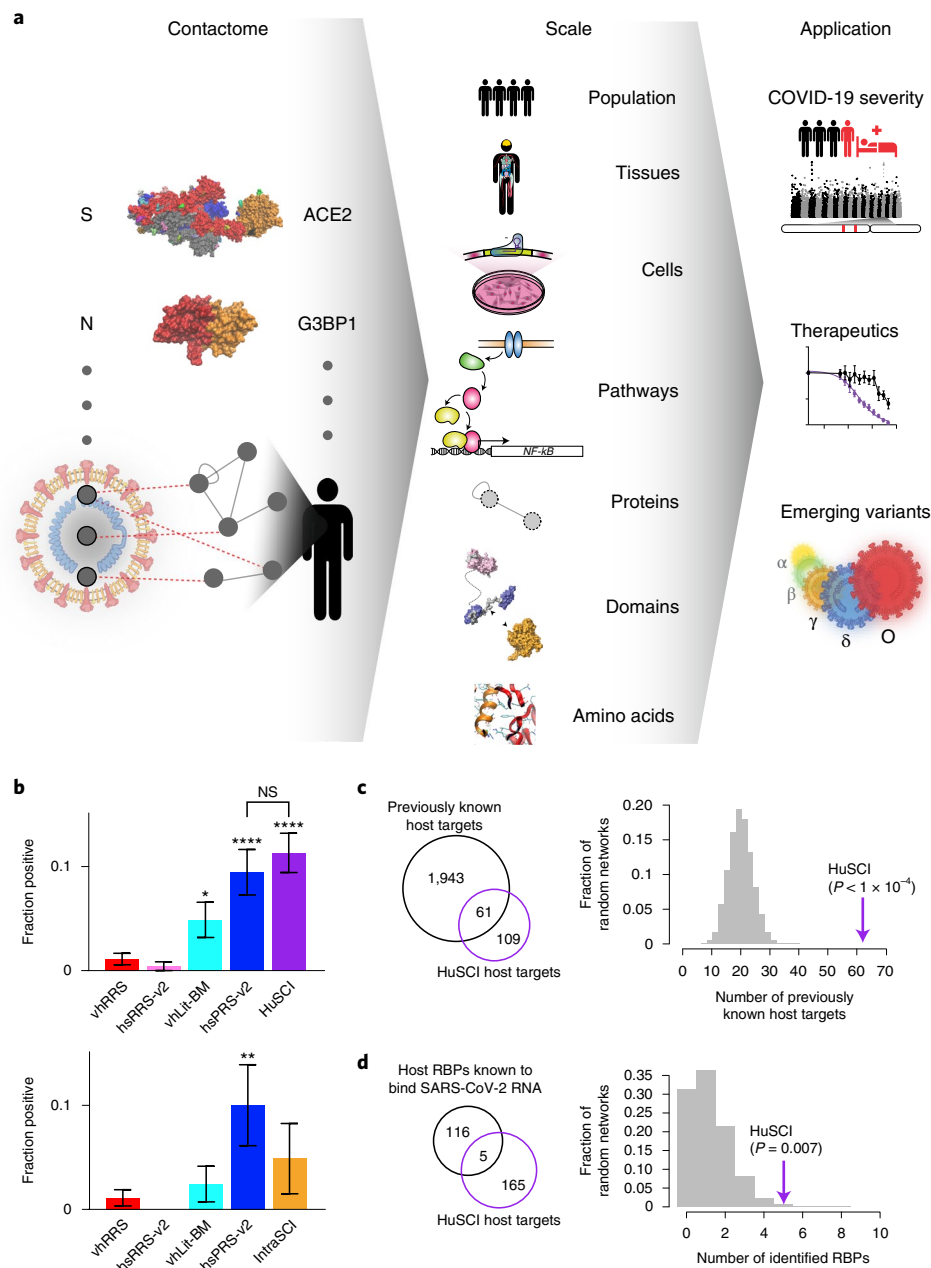
Fig. 1 | Generation and quality assessment of HuSCI. a, The contactome, the sum of physical contacts between viral and host macromolecules, mediates cellular perturbations that enable viral replication and cause disease manifestations. N, nucleocapsid protein; S, spike protein. **b,** Orthogonal validation; fraction of pairs that are yN2H-positive in HuSCI (top, $n = 282$ pair configurations representing 148 HuSCIs interaction pairs) and IntraSCI (bottom, $n = 41$ pair configurations for 25 IntraSCI interaction pairs), in the benchmark positive control sets hsPRS-v2 ($n = 180$ pair configurations for 60 interaction pairs) and vhLit-BM ($n = 164$ pair configurations for 40 interaction pairs), and negative control sets hsRRS-v2 ($n = 234$ pair configurations for 78 protein pairs) and vhRRS ($n = 360$ pair configurations for 178 protein pairs). Asterisks indicate significant differences from vhRRS benchmark (*$P = 0.023$; **$P = 0.005$; ****$P = 1.57 \times 10^{-5}$ hsPRS-v2, $P = 1.02 \times 10^{-7}$ HuSCI; NS, not significant; two-sided Fisher's exact test; center, proportion of positives; error bars, standard error of proportion). Precise P values for all dataset pairs, biological repeats and n for each test are shown in Supplementary Table 3. **c,** Overlap of SARS-CoV-2 targets identified in HuSCI with previously identified target proteins of other viruses (left) and actual overlap (arrow) compared to $n = 10,000$ randomized control networks (right) (one-sided, empirical $P < 0.0001$). **d,** Host targets identified in HuSCI overlap with RNA-binding proteins (RBPs) bound to SARS-CoV-2 RNA upon infection (left) and actual overlap (arrow) compared to $n = 10,000$ randomized control networks (right) (one-sided, empirical $P = 0.007$).

used two complementary assay versions (Extended Data Fig. 1a): (1) a plate-based version using 'bait' and 'prey' N-terminal fusion proteins encoded on low-copy plasmids and *GAL1-HIS3*-based growth selection (Y2H$_{HIS3}$)[2,3], and (2) a new system based on the Barcode Fusion Genetics (BFG)-Y2H technology[4], using a C-terminal fusion prey protein encoded from a high-copy plasmid and selecting cells expressing green fluorescent protein (*GAL1-GFP*) from a pooled liquid culture (Y2H$_{GFP}$)

(unpublished). Using Y2H$_{HIS3}$, 26 viral open reading frames (vORFs; Supplementary Table 1) were screened against 17,472 human ORFs (covering 83% of all pairings of human and viral protein-coding genes, that is 83% 'search space completeness') in both orientations; that is, as bait and prey (Extended Data Fig. 1a). Human candidate interactors were pairwise retested in triplicate against every vORF, yielding 118 interactions involving 14 viral and 92 human proteins. We refer to this

Y2H$_{HIS3}$-based human SARS-CoV-2 interactome dataset as HuSCI$_{HIS3}$. Using Y2H$_{GFP}$, 28 vORFs were screened against 14,627 human ORFs (70% completeness) (Extended Data Fig. 1a). After stringent filtering and *HIS3*-based verification, this yielded 93 interactions involving 13 viral and 84 human host proteins. We refer to this dataset as HuSCI$_{GFP}$ and to the union with HuSCI$_{HIS3}$ as HuSCI (Supplementary Table 1). We also carried out a targeted screen with previously identified SARS-CoV-1 host interactors; of the 62 testable orthologous SARS-CoV-2–human pairs, six were found to interact (HuSCI$_{ORTH}$) (Supplementary Table 2). Y2H$_{GFP}$ also yielded an intraviral SARS-CoV-2 interactome of 25 binary interactions among 19 vORFs (IntraSCI; Supplementary Table 1). Having collectively identified a contactome of 204 direct virus–host and 25 intraviral interactions among 170 host and 19 viral proteins, we next assessed data quality.

Seven interactions were identified in both HuSCI$_{GFP}$ and HuSCI$_{HIS3}$. Albeit nominally low, this overlap is consistent with the complementary nature of the assays and pipelines. Specifically, the screens interrogated incompletely overlapping protein sets and were each 50%–60% saturated. Each version used for screening has an assay sensitivity of 20%–25%[5] (fraction of detectable interactions); thus, the overlap is consistent with known screening parameters[2] and a low false-discovery rate. Moreover, from these parameters we can estimate that HuSCI covers 15%–22% of the complete contactome between SARS-CoV-2 and host proteins (Methods).

To further assess data quality experimentally, we compared detection rates of our datasets in the yeast-based nanoluciferase complementation assay (yN2H)[6] to those of established human positive and random reference sets (hsPRS-v2 and hsRRS-v2)[5,6]. As additional benchmarks, we derived a set of 55 well-documented binary interactions between human and coronavirus proteins from the curated literature (virus–host literature binary multiple reference set; vhLit-BM) and a virus–host random reference set (vhRRS) (Supplementary Table 3). We tested HuSCI, IntraSCI and each benchmark set by yN2H (Fig. 1b and Extended Data Fig. 1b). At a stringent scoring threshold of 1% vhRRS, the validation rates of both HuSCI alone and the union of HuSCI with IntraSCI (UnionSCI) were statistically indistinguishable from the two positive control sets (hsPRS-v2, $P = 0.76$; vhLit-BM, $P = 0.06$; Fisher's exact test versus UnionSCI), and each was significantly higher than those of the negative control sets (hsRRS-v2, $P = 4 \times 10^{-7}$; vhRRS, $P = 1 \times 10^{-7}$; Fisher's exact test versus UnionSCI; Fig. 1b and Supplementary Table 3). Thus, the biophysical quality of our virus–host contactome map is at least on par with high-quality interactions supported by multiple experiments in the curated literature. Although IntraSCI is too small for a separate evaluation by yN2H, 5 of 25 interactions overlap with a previous study[7] ($P = 4.6 \times 10^{-3}$, empirical test; Extended Data Fig. 1c).

The biological relevance of our virus–host contactome map is suggested by the observations that the identified host proteins are enriched for (1) known targets of other viruses[8] ($P < 1 \times 10^{-4}$, empirical test; Fig. 1c), (2) proteins that change phosphorylation status upon SARS-CoV-2 infection[9,10] ($P < 1 \times 10^{-4}$, empirical test; Extended Data Fig. 1d), (3) proteins that directly interact with SARS-CoV-2 RNA[11] ($P = 0.007$,

empirical test; Fig. 1d) and (4) proteins that change RNA-binding status during SARS-CoV-2 infection[11] ($P = 0.022$, empirical test; Extended Data Fig. 1e). These results demonstrate that IntraSCI and HuSCI (Fig. 2a) are of high biophysical quality and enriched for host proteins relevant to SARS-CoV-2 biology.

Complementarity of contactome and co-complex datasets
Previous studies investigating host and SARS-CoV-2 proteins used either affinity purification followed by mass spectrometry (AP-MS) to identify co-complex associations[9,12–15] or biotin identification (BioID) to find proteins in spatial proximity[16–18]. However, co-complex maps capture largely indirect associations in stable complexes that persist through affinity purification[2] and, likely due to experimental differences, the datasets exhibit limited agreement among each other (Extended Data Fig. 2a). For a subset of such co-complex associations, contacts can be computationally modeled[19]. In contrast, binary interactome maps provide direct contact partners and are enriched for regulatory interactions[2]. Despite these differences, 20 of the 204 HuSCI-interacting pairs were found in co-complex and BioID studies, and 58 (34%) of the 170 HuSCI host proteins were associated with a SARS-CoV-2 protein by these studies (Supplementary Table 1). Thus, the contactome map is consistent with previous indirect association datasets while providing substantial novelty.

Although SARS-CoV-2 primarily infects lung and airway tissues, it can spread to additional tissues and this expanded tropism is characteristic for COVID-19 and important for long COVID symptoms[20]. As previous SARS-CoV-2 interaction datasets could only detect host proteins expressed in the specific assay cell lines, we wondered whether HuSCI was also complementary in terms of the tissue specificity of identified host proteins. Using the Human Protein Atlas (HPA)[21], we defined 'tissue-specific' and 'common' human proteins (Supplementary Table 4). Whereas the AP-MS and BioID data are biased toward common host proteins, HuSCI is more representative of the human proteome and shows good coverage of proteins expressed in the diverse tissues in which SARS-CoV-2 RNA has been detected[22] (Fig. 2b, Extended Data Fig. 2b,c and Supplementary Table 4). Thus, the HuSCI contactome has unique advantages for understanding tissue-specific perturbations by SARS-CoV-2.

SARS-CoV-2 targeted functions
To understand which host functions are directly perturbed by the virus, we investigated SARS-CoV-2 targeted pathways. Broad functions enriched among host proteins include (1) immune response, (2) viral process, (3) protein ubiquitination, (4) cytoskeleton and (5) vesicle-mediated transport (Fig. 2c). These largely agree with functions identified in association and proximity datasets[9,12–18] (Supplementary Table 5). Focusing on immune pathways, we noticed that NSP9, NSP14 and NSP16 contact key regulators of cytokine production such as REL (c-REL proto-oncogene, NF-κB subunit), IKBKG (inhibitor of NF-κB kinase regulatory subunit gamma, also known as IKKγ or NEMO) and TRAF2 (tumor necrosis factor (TNF) receptor-associated factor 2). HuSCI interactors of the membrane-spanning NSP6 were enriched
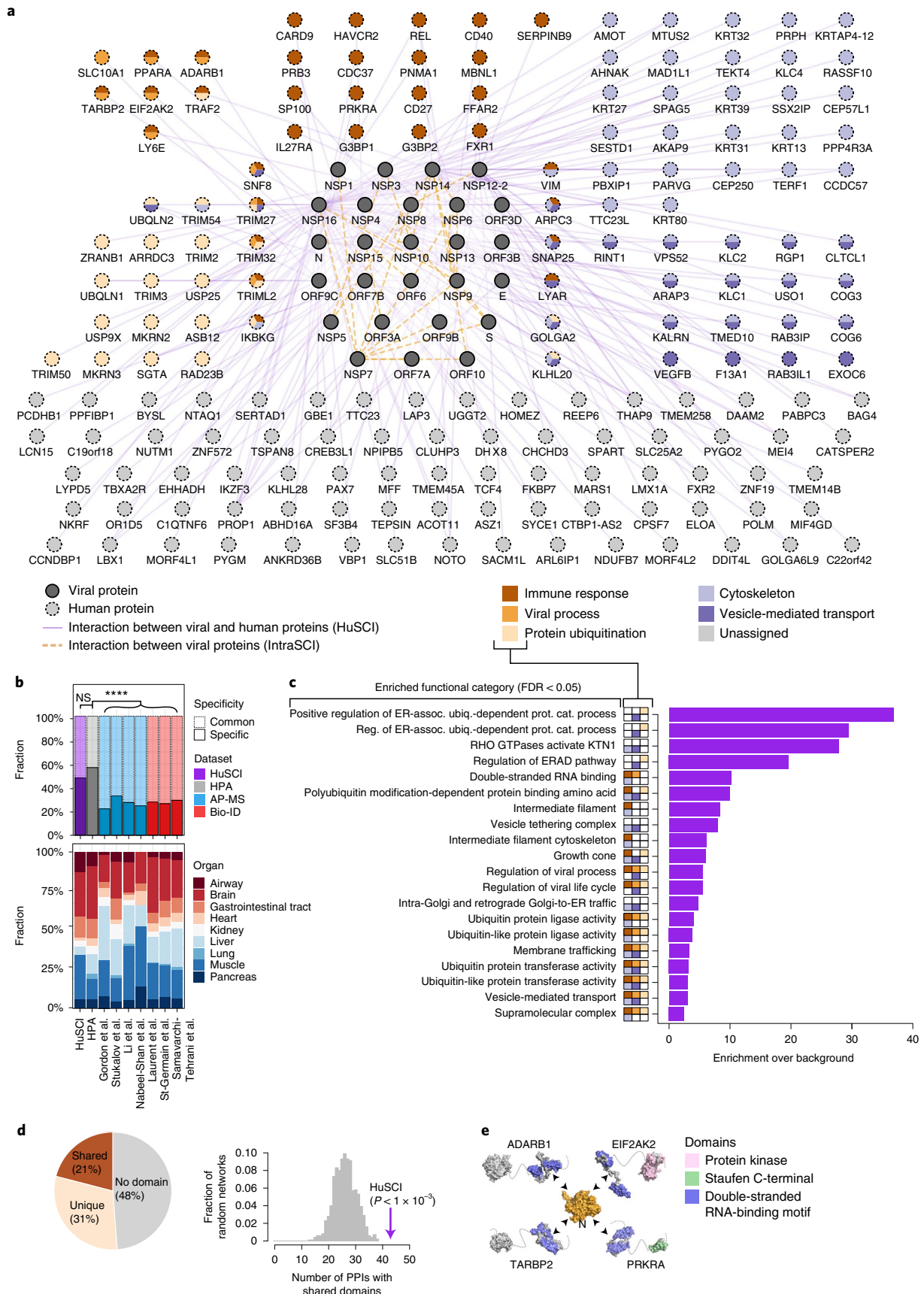
Fig. 2 | Network representation and functional assessment of HuSCI.
a, Combined HuSCI and IntraSCI networks. Node colors of human proteins represent broad enriched functions as indicated in legend. Node labels for human proteins correspond to approved HGNC symbols; accession identifiers and descriptions are listed in Supplementary Table 1. **b**, Proportion of host targets in common and specific expression groups in all (top) and in SARS-CoV-2 RNA-positive organs (bottom) across eight datasets: purple, HuSCI; gray, HPA[21]; blue, AP-MS datasets from Gordon et al.[12,13], Stukalov et al.[9], Li et al.[14] and Nabeel-Shah et al.[15]; red, BioID datasets from Laurent et al.[16], St-Germain et al.[17] and Samavarchi-Tehrani et al.[18]. Two-sided Fisher's exact test, Bonferroni adjusted $P < 0.0001$. Full statistical details and exact P values are listed in Supplementary Table 4. **c**, Functions enriched among host proteins found in HuSCI ($P = 0.05$, Fisher's exact test with FDR correction). Broad functional groups are indicated

in small boxes according to legend in panel a. Full statistical details are listed in Supplementary Table 5. ER-assoc. ubiq.-dependent prot. cat., Endoplasmic reticulum-associated ubiquitin-dependent protein catabolic. **d**, Proportion of virus–host interactions in which the human protein has domains that are present in other interactors of the viral protein (shared), not present in other interactors of the viral protein (unique) or no annotated domains (left) and number of shared-domain interactions in HuSCI (arrow) compared to $n = 1,000$ randomized control networks (gray distribution) (right). One-sided empirical $P < 0.001$. PPI, protein-protein interactions. **e**, Exemplary 'shared-domain interaction' between the viral nucleocapsid protein and four interactors containing a double-stranded RNA-binding motif. Domain colors according to legend; gray parts lack domain annotations.

for immune receptors ($P < 0.01$, empirical test), including CD40 and IL27RA (IL27 receptor subunit alpha). Intriguingly, NSP6 also directly interacts with LY6E, a host restriction factor that limits viral entry

for SARS-CoV-2 and other coronaviruses[23]. Several other targets are RNA-binding proteins that function in innate immunity and response to viral infection[24]. MKRN2, together with G3BP1/2, has been suggested

a

- ⬤ Viral protein
- ⬤ Human protein
- — Interaction between viral and human proteins (HuSCI)
- ╌╌ Interaction between viral proteins (IntraSCI)

- ■ Immune response
- ■ Viral process
- ■ Protein ubiquitination
- ■ Cytoskeleton
- ■ Vesicle-mediated transport
- ■ Unassigned

b Specificity: Common, Specific. Dataset: HuSCI, HPA, AP-MS, Bio-ID. Organ: Airway, Brain, Gastrointestinal tract, Heart, Kidney, Liver, Lung, Muscle, Pancreas.

c Enriched functional category (FDR < 0.05)

- Positive regulation of ER-assoc. ubiq.-dependent prot. cat. process
- Reg. of ER-assoc. ubiq.-dependent prot. cat. process
- RHO GTPases activate KTN1
- Regulation of ERAD pathway
- Double-stranded RNA binding
- Polyubiquitin modification-dependent protein binding amino acid
- Intermediate filament
- Vesicle tethering complex
- Intermediate filament cytoskeleton
- Growth cone
- Regulation of viral process
- Regulation of viral life cycle
- Intra-Golgi and retrograde Golgi-to-ER traffic
- Ubiquitin protein ligase activity
- Ubiquitin-like protein ligase activity
- Membrane trafficking
- Ubiquitin protein transferase activity
- Ubiquitin-like protein transferase activity
- Vesicle-mediated transport
- Supramolecular complex

Enrichment over background

d Shared (21%), No domain (48%), Unique (31%). Number of PPIs with shared domains. HuSCI ($P < 1 \times 10^{-3}$)

e ADARB1, EIF2AK2, TARBP2, PRKRA. Domains: Protein kinase, Staufen C-terminal, Double-stranded RNA-binding motif.

to regulate olfactory signaling mRNAs[25], pointing to potential mechanistic links underlying anosmia in COVID-19. Thus, direct SARS-CoV-2 protein interactors function in immune pathways and viral processes relevant to COVID-19.

Viral proteins contact shared host-protein domains

The restricted size of viral genomes limits their coding potential. We therefore wondered to what extent this limitation yielded viral proteins that bind multiple human proteins via target-shared domains, thus offering opportunities for structure-based drug discovery. We sought domains shared by multiple human targets of each viral protein. In the contactome, SARS-CoV-2 proteins engaged in 43 interactions involving such shared domains (21% of HuSCI; $P < 0.001$, empirical test; Fig. 2d, Extended Data Fig. 2d and Supplementary Table 6), corresponding to 22 significant virus protein-domain pairs ($P < 0.001$, empirical test; Supplementary Table 6). Although the difference was not statistically significant, the 21% proportion of the virus–host contactome with shared-domain interactions was numerically higher than the corresponding 17% in the human reference interactome network (HuRI)[26]. Specific examples in HuSCI include four interactors of the nucleocapsid protein sharing the double-stranded RNA-binding motif ($P < 0.05$, Fisher's test; Fig. 2e) and the recently confirmed finding that viral nucleocapsid protein interacts with the NTF2 domains of G3BP1 and G3BP2[27]. Disease-causing mutations are located in the interaction interfaces of the enriched domains of several human proteins (for example, TNF receptor domain of CD40 (ref. [28]) or zf-CCCH in MKRN3 (ref. [29])). Thus, recurrent structural themes may reflect binding mechanisms that are subject to modulation by human coding variants affecting infection outcome[30,31] or by rationally designed drugs.

HuSCI links to COVID-19 risk loci

The severity of COVID-19 symptoms and outcomes are highly variable, and understanding the underlying molecular mechanisms may enable effective treatments. Recently, two independent meta-studies identified genetic loci that are associated with severe COVID-19 illness[32,33] (Fig. 3a and Extended Data Fig. 3a), but mechanistic links to viral infection remain unknown. Similarly, several preconditions increase the risk of severe COVID-19, but for these, the molecular links are also poorly understood. At least two models can help to conceptualize how this genetic variation relates to virally targeted host proteins. In a 'direct' model, genetic variation in targeted host proteins modulates disease outcome, exemplified by the interaction of adenovirus E1A oncoprotein with the tumor suppressor protein pRb[34]. In an alternative 'indirect' model, genetic variation in the network neighborhood of targeted host proteins modulates the downstream effects and thereby influences disease outcome. A precedent for this model was observed in a plant system, where pathogen-targeted host proteins tend to interact with proteins relevant to disease severity and fitness (encoded by highly variable genes under balancing selection)[35]. The availability of a high-quality contactome map enabled us to address this fundamental question for COVID-19. Because bias toward well-studied proteins in the SARS-CoV-2 literature[36] (Fig. 3b and Extended Data Fig. 3b) limits mechanistic understanding and can cause artifacts, we focused our analyses on systematic protein interaction datasets. The direct model was not supported, given that no targeted host protein from HuSCI was encoded from a critical illness associated locus[32,33] ('critical illness proteins'), and only one (HLA-G, associated with ORF3) was found in a single co-complex study[9]. Investigating the indirect model, we sought contacts between targeted host proteins and critical illness proteins, finding 20 ($P = 0.002$, empirical test; Fig. 3c)[32] and 8 ($P = 0.012$, empirical test; Extended Data Fig. 3a, c)[33] in the binary HuRI host network map. In contrast, the virus-associated host-protein sets from AP-MS studies[9,12,13] interact with no more critical illness proteins than expected by chance (Extended Data Fig. 3d). Functionally, the HuSCI host-target proteins linking critical illness to SARS-CoV-2 proteins are enriched

in microtubule organization, membrane trafficking and TNF signaling annotations (Supplementary Table 7). Intriguingly, three of seven direct OAS1 interactors are targeted by NSP14 and NSP16, and all three have Golgi- and membrane trafficking-related functions, providing protein contacts that support the finding that the Neanderthal-derived protective OAS1 variant promotes degradation of viral RNA in endoplasmic reticulum- and Golgi-derived virus replication organelles[37]. These observations indicate that, consistent with the indirect model, clinically relevant genetic variation acts in the local network neighborhood of viral contact proteins.

To further explore the local subnetworks surrounding targeted host proteins and their links to human genetic variation, we identified 204 subnetwork communities in HuRI[26] (Fig. 3d) that were significantly targeted by SARS-CoV-2 (nominal $P < 0.05$, Fisher's exact test; Supplementary Table 8). Examples include community 28, enriched for 'negative regulation of viral transcription' (false discovery rate (FDR) = 0.0018; Fig. 3d) and community 52, enriched for 'Arp2/3 complex-mediated actin nucleation' (FDR = 0.0002; Supplementary Table 8). The Arp2/3 complex enables human respiratory viruses to spread among adjacent cells without forming virions[38], and ARPC3 scored among the top 50 in two CRISPR screens for SARS-CoV-2 host factors[39,40]. We then asked whether direct viral target proteins and proteins in each community are encoded by genes associated with human traits of 114 uniformly processed genome-wide association studies (GWASs)[41]. Variation in genes encoding direct viral targets was only associated with 'depression' (FDR = 0.03, MAGMA). In contrast, among the communities, genetic variation associated with severe COVID-19 illness was associated with ten virus-targeted communities, more communities than any other human trait. In contrast, host-protein sets from AP-MS studies were enriched in fewer communities (nominal $P < 0.05$, Fisher's exact test; Extended Data Fig. 3e and Supplementary Table 8), and only one host-protein-enriched community each from two AP-MS datasets was enriched for genetic variation associated with severe COVID-19 (refs. [13,14]) (Li et al. community 14 and Gordon et al. community 11; Extended Data Fig. 3f). Intriguingly, of the 14 human traits (from 15 studies) associated with 20 additional HuSCI-target-enriched communities, 8 traits (from 9 studies) are clinical risk factors for severe COVID-19 and long COVID, including high body mass index (BMI)[42], hypothyroidism[43] and schizophrenia[44] ($P = 0.01$, Fisher's exact test; Fig. 3d, Extended Data Fig. 3e,f and Supplementary Table 8). These links between viral targets and genetic variation associated with COVID-19 comorbidities open the possibility that this genetic variation may impact the course of infection and severity of COVID-19 independent of trait manifestation. Other traits associated with host-target-enriched communities, such as neuroticism, have not been linked to COVID-19 symptoms, possibly because the genetic influence is masked by confounding parameters such as behavior[45], and should be considered in the future. Together, these results suggest that the HuSCI contactome map is a powerful and unique resource for studying molecular mechanisms by which human genetics affect the outcome of SARS-CoV-2 infection.

Validation of pathways and host targets

We next explored specific hypotheses for viral proteins and human target functions. Both literature reports and our analyses suggest a role for NF-κB immune signaling in SARS-CoV-2 infection. Because we observed multiple interactions of viral proteins with different members of the NF-κB signaling pathway, we used reporter assays to determine whether and in which direction (that is, activating or inhibiting) viral factors modulate pathway activity. Transfection of NSP14, which interacts with multiple positive NF-κB regulators, resulted in dose-dependent transcriptional activation of NF-κB and even further augmented NF-κB activity following proinflammatory TNF-α stimulation in HEK293 cells (Fig. 4a,b, Extended Data Fig. 4a,b and Supplementary Table 9). This finding suggests that SARS-CoV-2 can induce a proinflammatory state during COVID-19 via direct interaction of NSP14 with NF-κB activators.

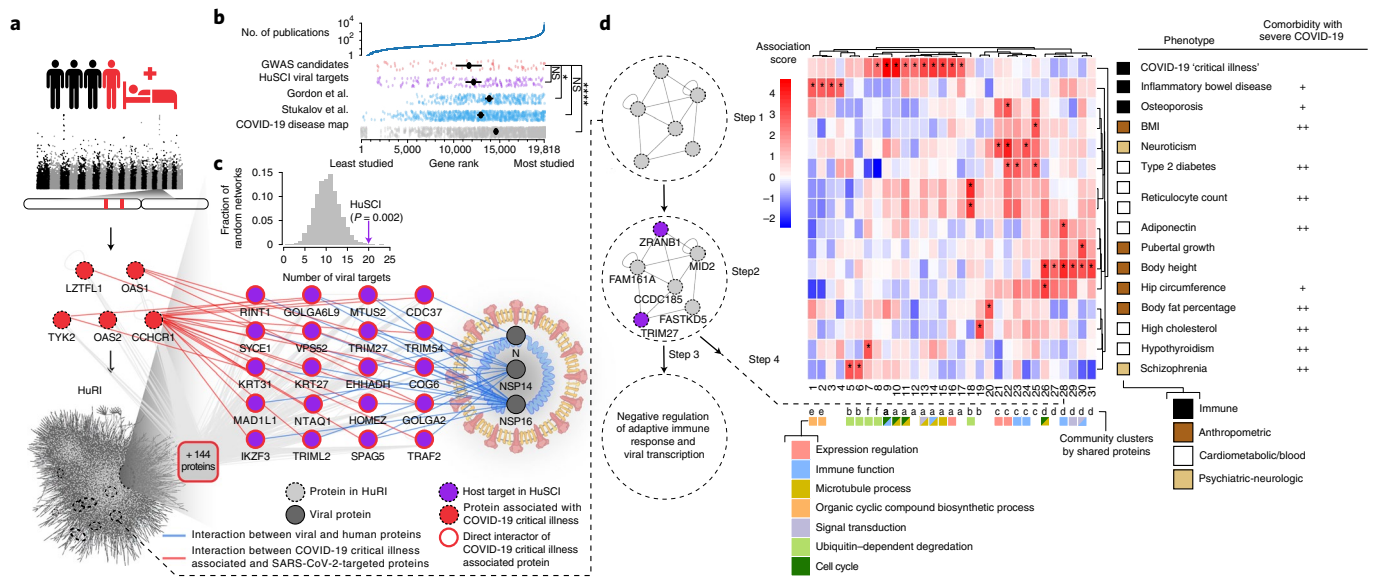**Fig. 3 | HuSCI host targets link to genetic variation for severe COVID-19.**
a, HuRI interactors (gray nodes) of COVID-19 'critical illness proteins' loci (seed, red nodes) include a significant ($P = 0.0009$, empirical test) number of direct SARS-CoV-2 targeted proteins (purple nodes). A total of 144 additional seed protein interactors are not resolved individually. Node and edge colors according to legend. **b**, Genes in indicated COVID-19 datasets ranked across the human genome by number of publications. Number of publications is indicated by the top panel on log scale. Asterisks indicate significant differences relative to COVID-19-associated genes[32,33]; NS, $P = 1$ (HuSCI) and $P = 0.36$ (Stuckalov et al.); *$P = 0.047$; ****$P = 0.000014$, two-sided Mann–Whitney U test, Bonferroni correction, from top to bottom: $n = 45, 170, 383, 876$ and 849; error bars are 95% confidence intervals of the mean, calculated by 1,000 bootstrap samples. **c**, Virus-interactor enrichment: number of direct SARS-CoV-2 protein interacting HuSCI proteins in the subnetwork formed by proteins encoded by seed proteins[32,33] and their first-level interactors (arrow) compared to $n = 10,000$ randomized control networks (gray distribution). **d**, Of 3,603 communities in Human Reference Interactome (HuRI) with ≥4 members (step 1), 204 are significantly targeted by SARS-CoV-2 (two-sided nominal $P < 0.05$; Fisher's exact test) (step 2); Gene Ontology (GO) enrichment identifies functions associated with each community (step 3); and MAGMA identifies 31 communities significantly associated with human traits (FDR < 0.05) (step 4), the great majority of which are COVID-19 comorbidities. Example community 28 is significantly targeted by SARS-CoV-2 in HuSCI (two-sided $P = 0.0078$; Fisher's exact test, uncorrected) and enriched for negative regulation of adaptive immune response and viral transcription. Functional descriptors in squared boxes according to legend (Supplementary Table 8); relation of indicated traits to COVID-19 is indicated in rightmost column as general link (+) (e.g., via immunity) and clinical evidence for modulation of diseases symptoms and risk for severe or long COVID (+ + ; Extended Data Fig. 3f). BMI, body mass index.

These results are corroborated by a study that implicates IMPDH2 in NF-κB pathway activation by NSP14 (ref. [46]). Moreover, transcriptional profiling experiments have demonstrated NF-κB activation in HEK293 cells and in patients following SARS-CoV-2 infection[47,48]. As TNF-α has a central role in the cytokine storm that contributes to many COVID-19 deaths[49], the observation that SARS-CoV-2 activates this system in a cell-intrinsic manner may have therapeutic implications.

We explored the role of the NSP14 interactor IKBKG/NEMO, an essential mediator of canonical NF-κB signaling[50], for transcriptional activation. We generated IKBKG HEK293 knockout (KO) clones (Extended Data Fig. 4) and checked for NF-κB activation in three independent clones after NSP14 transfection (Fig. 4c). IKBKG deficiency abolished NF-κB activation in response to TNF-α and severely impaired NSP14-induced NF-κB activation, providing evidence for a functional role of IKBKG in driving NF-κB activation by NSP14. Interestingly, the residual NF-κB reporter induction upon NSP14 expression in the KO cells indicates that other NSP14 interactors (for example, TRAF2 and REL) contribute to the full NF-κB transcriptional response.

We wondered whether NF-κB signaling proteins and virally targeted host proteins in enriched functional groups other than 'immune response' (Fig. 2a) are important for viral replication. After generating A549 alveolar basal epithelial adenocarcinoma cells that exogenously express human ACE2 (A549-ACE2), we quantified viral replication in the presence and absence of CRISPR-Cas9-mediated KO of viral-target-encoding genes. Of eight genes that were selected from enriched functional groups and successfully knocked out, deletion of five (63%) resulted in a significant decrease of viral replication (Fig. 4d). Intriguingly, deletion of three NSP14-interacting proteins of

the NF-κB signaling system (REL, IKBKG and TRAF2) resulted in strong reduction of viral replication (Fig. 4d and Extended Data Fig. 4f,g). This finding is consistent with a model in which SARS-CoV-2 directly activates NF-κB via NSP14, with this activation being required for successful viral replication. Deletion of kinesin light chain 1 (KLC1), a cargo adaptor protein for microtubule mediated transport, caused reduction of replication by ~80% ($P < 0.0001$, Kruskal–Wallis test). Beyond this observation, deletion of ubiquitin-specific peptidase 25 (USP25), which has antiviral functions in influenza and herpes infections[51], resulted in essentially complete elimination of viral replication without impacting cell growth, suggesting that human USP25 is required by SARS-CoV-2 (Fig. 4d, Extended Data Fig. 4f,g and Supplementary Table 10).

Inspired by the strong effect on viral replication, we explored USP25 as an antiviral drug target using the small molecule AZ1, which effectively inhibits USP25 and USP28 enzymatic activity[52]. Using an infectious clone-derived SARS-CoV-2 (icSARS-COV-2) harboring a mNeonGreen marker[53], we showed that treatment with 10 μM AZ1 effectively inhibits SARS-CoV-2 replication in Vero E6 cells (Fig. 4e). Next, we used an independent icSARS-CoV-2 expressing nanoluciferase[54] for dose titrations. The AZ1 compound interfered with SARS-CoV-2 replication with half-maximum effective concentration (EC_{50}) values of 0.8 μM and 0.1 μM in HEK293-ACE2 and Vero E6 cells, respectively (Fig. 4f and Supplementary Table 11), on par with the effects of the clinically approved remdesivir (Extended Data Fig. 4h). Effective concentrations are in the range of the half-maximal inhibitory concentration determined for inhibition of USP25/28 enzymatic activities[52], further supporting that USP25 is necessary for SARS-CoV-2 replication. Although the antiviral activity of AZ1 was independently identified in
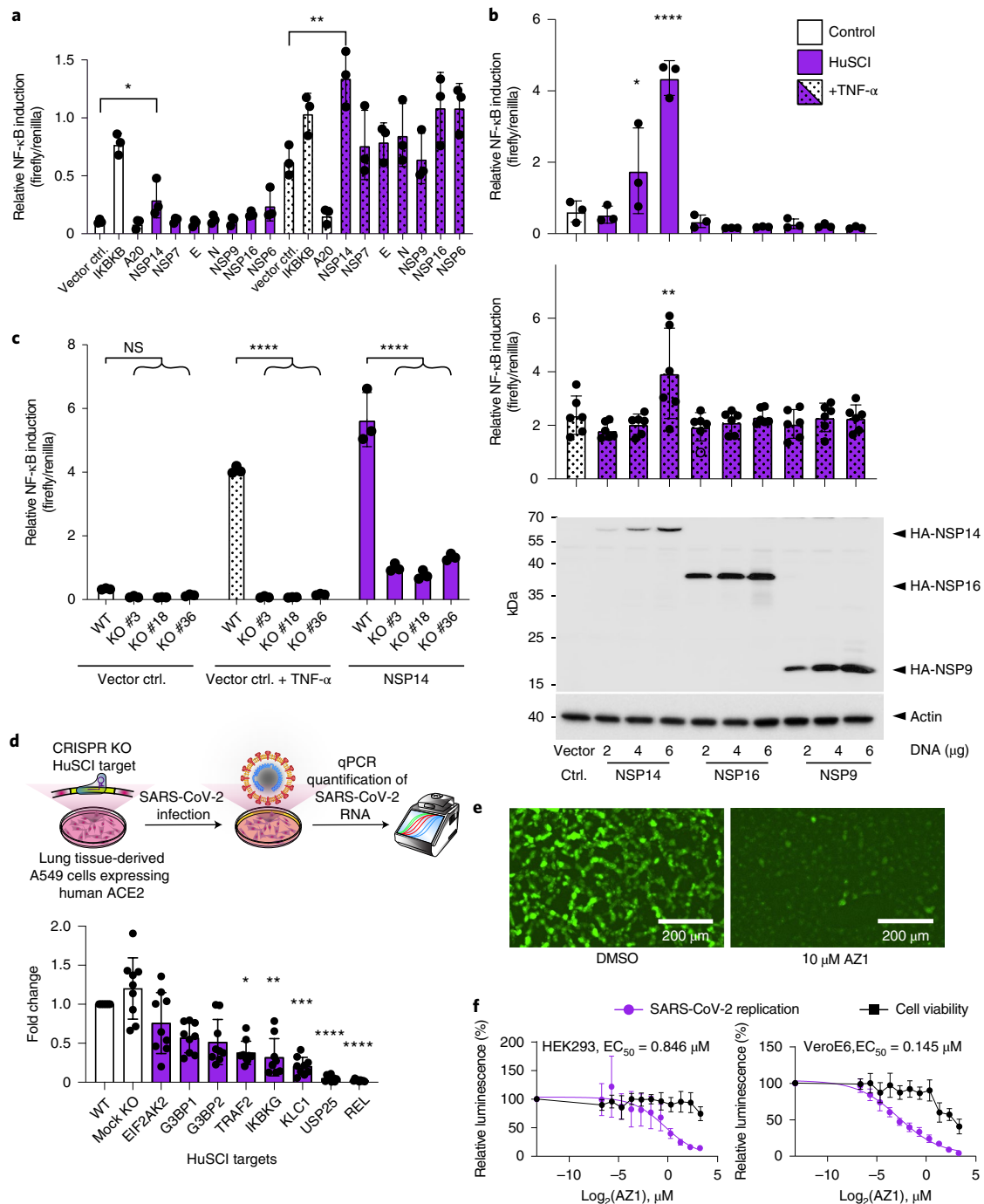
Fig. 4 | Validation of pathways and host targets. a, Relative NF-κB transcriptional reporter activity in unstimulated (left) and TNF-α-stimulated conditions (one-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test, $P = 0.0395$ and $P = 0.0047$, respectively). Error bars represent standard deviation of the mean, $n = 3$. **b**, Relative NF-κB transcriptional reporter activity at different amounts of transfected viral protein-encoding plasmid in unstimulated (top) and TNF-α-stimulated conditions (middle) (one-way ANOVA with Dunnett's multiple comparisons test: $*P = 0.0183$, $****P < 0.0001$ and $**P = 0.0012$, respectively). Error bars represent standard deviation of the mean, $n = 3$ (top) and $n = 6$ (middle). Representative anti-hemagglutinin (HA) western blot demonstrating levels of tagged viral protein in titration experiments relative to actin beta (ACTB) loading controls (bottom). **a** and **b**, Precise P values, biological repeats and n for each test are shown in Extended Data Fig. 4 and Supplementary Table 9. **c**, Relative NF-κB transcriptional reporter activity under unstimulated (left), TNF-α-stimulated (middle) and NSP14-induced conditions in wild-type (WT) and three independent IKBKG KO clones of HEK293 cells (two-

way ANOVA with Dunnett's multiple comparisons test). Error bars represent standard deviation of the mean, $n = 3$. Precise P values, biological repeats and n for each test are shown in Extended Data Fig. 4. ctrl., control. **d**, Schematic of viral replication assay (top) and viral replication in wild-type, mock KO and CRISPR KOs of the indicated HuSCI host targets (bottom) (Kruskal–Wallis with Dunn's multiple comparisons test, $* P = 0.031$, $** P = 0.0047$, $*** P = 0.0003$, $**** P < 0.0001$, respectively). Error bars represent standard deviation of the mean, $n = 9$. Precise P values, biological repeats and n for each test are shown in Extended Data Fig. 4 and Supplementary Table 10. **e**, Fluorescence microscopy images showing replication of icSARS-CoV-2-mNeonGreen in infected Vero E6 cells treated with 10 μM AZ1 or solvent (DMSO, dimethylsulfoxide). **f**, Cell viability and relative replication of icSARS-CoV-2-nanoluciferase in HEK293 cells (left) and Vero E6 cells (right) at different concentrations of AZ1. The EC_{50} values were calculated with a variable slope model. Error bars represent standard deviation of the mean, $n = 8$ biological repeats and full analysis in Supplementary Table 11.

a small-molecule screen[55], our results inform mechanistic studies by identifying NSP16 as a viral interaction partner. NSP16 and associated complexes methylate viral RNA to prevent its detection and destruction by the innate immune system[56,57]. The stable recruitment of USP25 may protect this complex from ubiquitination and degradation by the host defense machinery. Although elucidating precise mechanisms will require further studies, these findings illustrate the high potential of the HuSCI contactome map in helping to understand and inhibit the SARS-CoV-2 life cycle.

Perturbed contactome in SARS-CoV-2 variants

Evaluating the impact of novel viral strains on the contactome has been largely restricted to spike protein interactions with ACE2 and antibodies[58]. Wondering if coding variants in other viral proteins perturb the contactome and thereby modulate viral effects, we explored the potential of 19 SARS-CoV-2 mutations in 14 variants of 9 proteins from the Alpha, Beta, Gamma and Delta strains to alter interactions with host contact targets in HuSCI (Supplementary Table 12). Indeed, some mutations resulted in perturbed interactions. The Alpha strain mutant combination D3L, S235F in the nucleocapsid protein reduced interaction with ARPC3, the SARS-CoV-2 host factor discussed above. Similarly, the Beta-strain mutation P71L in the envelope (E) protein diminished the interaction with BAG4, an antiapoptotic protein involved in TNF signaling (Extended Data Fig. 5). Although it is currently unknown whether the respective interactions promote viral replication or facilitate immune recognition, the observed changes demonstrate the plasticity of the contactome and, together with recent reports of increased replication of the Delta strain[59], strongly suggest that this dimension of viral evolution should also be monitored to assess the risk posed by emerging variants.

Discussion

In summary, we present a validated contactome map, HuSCI, which provides direct interactions between SARS-CoV-2 and human target proteins in pathways and tissues relevant to COVID-19. HuSCI enables identification of paths of direct contact between viral target proteins and proteins encoded from loci that modify the risk for critical COVID-19 illness and important comorbidities. Examining specific hypotheses for both viral and host proteins, we demonstrate that NSP14 activates the NF-κB pathway even beyond pathway activation by cytokines. Moreover, the majority of the virally targeted host proteins we evaluated, including key NF-κB regulators, are required for efficient SARS-CoV-2 replication. For one of these targeted host proteins, USP25, we confirm that a small-molecule inhibitor can dramatically reduce viral replication and implicate a mechanism for this potential therapeutic. Last, we demonstrate that coding changes in SARS-CoV-2 strains perturb the intracellular interactome. We anticipate that these findings and the contactome resource will stimulate important research toward characterizing new viral strains, understanding the mechanism of COVID-19 symptoms and developing therapies for current and future pandemics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-022-01475-z.

References

1. Nalbandian, A. et al. Post-acute COVID-19 syndrome. *Nat. Med.* **27**, 601–615 (2021).
2. Yu, H. et al. High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
3. Altmann, M. et al. Extensive signal integration by the phytohormone protein network. *Nature* **583**, 271–276 (2020).
4. Yachie, N. et al. Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* **12**, 863 (2016).
5. Braun, P. et al. An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6**, 91–97 (2009).
6. Choi, S. G. et al. Maximizing binary interactome mapping with a minimal number of assays. *Nat. Commun.* **10**, 3907 (2019).
7. Li, Y. et al. SARS-CoV-2 induces double-stranded RNA-mediated innate immune responses in respiratory epithelial-derived cells and cardiomyocytes. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2022643118 (2021).
8. Orchard, S. et al. The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **42**, D358–D363 (2014).
9. Stukalov, A. et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **594**, 246–252 (2021).
10. Bouhaddou, M. et al. The global phosphorylation landscape of SARS-CoV-2 infection. *Cell* **182**, 685–712.e19 (2020).
11. Kamel, W. et al. Global analysis of protein-RNA interactions in SARS-CoV-2 infected cells reveals key regulators of infection. *Mol. Cell* **81**, 2851–2867 (2021).
12. Gordon, D. E. et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, eabe9403 (2020b).
13. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020a).
14. Li, J. et al. Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med (N Y)* **2**, 99–112.e7 (2021).
15. Nabeel-Shah, S. et al. SARS-CoV-2 nucleocapsid protein binds host mRNAs and attenuates stress granules to impair host stress response. *iScience* **25**, 103562 (2022).
16. Laurent, E. M. N. 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. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.28.272955 (2020).
17. St-Germain, J. R. et al. A SARS-CoV-2 BioID-based virus-host membrane protein interactome and virus peptide compendium: new proteomics resources for COVID-19 research. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.28.269175 (2020).
18. Samavarchi-Tehrani, P. et al. A SARS-CoV-2–host proximity interactome. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.03.282103 (2020).
19. Wierbowski, S. D. et al. A 3D structural SARS-CoV-2–human interactome to explore genetic and drug perturbations. *Nat. Methods* **18**, 1477–1488 (2021).
20. Callard, F. & Perego, E. How and why patients made long covid. *Soc. Sci. Med.* **268**, 113426 (2021).
21. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
22. Dorward, D. A. et al. Tissue-specific immunopathology in fatal COVID-19. *Am. J. Respir. Crit. Care Med.* **203**, 192–201 (2021).
23. Zhao, X. et al. LY6E restricts entry of human coronaviruses, including currently pandemic SARS-CoV-2. *J. Virol.* **94**, e00562-20 (2020).
24. Garcia-Moreno, M. et al. System-wide profiling of RNA-binding proteins uncovers key regulators of virus infection. *Mol. Cell* **74**, 196–211 (2019).
25. Zanzoni, A., Spinelli, L., Ribeiro, D. M., Tartaglia, G. G. & Brun, C. Post-transcriptional regulatory patterns revealed by protein-RNA interactions. *Sci. Rep.* **9**, 4302 (2019).
26. Luck, K. et al. A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).

27. Kruse, T. et al. Large scale discovery of coronavirus-host factor protein interaction motifs reveals SARS-CoV-2 specific mechanisms and vulnerabilities. *Nat. Commun.* **12**, 6761 (2021).

28. Ferrari, S. et al. Mutations of CD40 gene cause an autosomal recessive form of immunodeficiency with hyper IgM. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12614–12619 (2001).

29. de Vries, L., Gat-Yablonski, G., Dror, N., Singer, A. & Phillip, M. A novel MKRN3 missense mutation causing familial precocious puberty. *Hum. Reprod.* **29**, 2838–2843 (2014).

30. Zhong, Q. et al. An inter-species protein-protein interaction network across vast evolutionary distance. *Mol. Syst. Biol.* **12**, 865 (2016).

31. Sahni, N. et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647–660 (2015).

32. Pairo-Castineira, E. et al. Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).

33. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).

34. Whyte, P. et al. Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. *Nature* **334**, 124–129 (1988).

35. Weßling, R. et al. Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* **16**, 364–375 (2014).

36. Ostaszewski, M. et al. COVID19 Disease Map, a computational knowledge repository of virus-host interaction mechanisms. *Mol. Syst. Biol.* **17**, e10387 (2021).

37. Soveg, F. W. et al. Endomembrane targeting of human OAS1 p46 augments antiviral activity. *eLife* **10**, e71047 (2021).

38. Cifuentes-Muñoz, N., Dutch, R. E. & Cattaneo, R. Direct cell-to-cell transmission of respiratory viruses: the fast lanes. *PLoS Pathog* **14**, e1007015 (2018).

39. Zhu, Y. et al. A genome-wide CRISPR screen identifies host factors that regulate SARS-CoV-2 entry. *Nat. Commun.* **12**, 961 (2021).

40. Daniloski, Z. et al. Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* **184**, 92–105.e16 (2021).

41. Barbeira, A. N. et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol* **22**, 49 (2021).

42. Bliddal, S. et al. Acute and persistent symptoms in non-hospitalized PCR-confirmed COVID-19 patients. *Sci. Rep.* **11**, 13153 (2021).

43. Whiting, A., Reyes, J. V. M., Ahmad, S. & Lieber, J. Post-COVID-19 fatigue: a case of infectious hypothyroidism. *Cureus* **13**, e14815 (2021).

44. Mohan, M., Perry, B. I., Saravanan, P. & Singh, S. P. COVID-19 in people with schizophrenia: potential mechanisms linking schizophrenia to poor prognosis. *Front. Psychiatry* **12**, 666067 (2021).

45. VanderWeele, T. J. Genetic self knowledge and the future of epidemiologic confounding. *Am. J. Hum. Genet.* **87**, 168–172 (2010).

46. Li, T. et al. SARS-CoV-2 Nsp14 activates NF-κB signaling and induces IL-8 upregulation. Preprint at *bioRxiv* https://doi.org/10.1101/2021.05.26.445787 (2021).

47. Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).

48. Sun, G. et al. Comparative transcriptomic analysis of SARS-CoV-2 infected cell model systems reveals differential innate immune responses. *Sci. Rep.* **11**, 17146 (2021).

49. Costela-Ruiz, V. J., Illescas-Montes, R., Puerta-Puerta, J. M., Ruiz, C. & Melguizo-Rodríguez, L. SARS-CoV-2 infection: the role of cytokines in COVID-19 disease. *Cytokine Growth Factor Rev.* **54**, 62–75 (2020).

50. Hayden, M. S. & Ghosh, S. Regulation of NF-κB by TNF family cytokines. *Semin. Immunol.* **26**, 253–266 (2014).

51. Lin, D. et al. Induction of USP25 by viral infection promotes innate antiviral responses by mediating the stabilization of TRAF3 and TRAF6. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11324–11329 (2015).

52. Wrigley, J. D. et al. Identification and characterization of dual inhibitors of the USP25/28 deubiquitinating enzyme subfamily. *ACS Chem. Biol.* **12**, 3113–3125 (2017).

53. Xie, X. et al. An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe* **27**, 841–848.e3 (2020).

54. Hou, Y. J. et al. SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* **182**, 429–446 (2020).

55. Grodzki, M. et al. Genome-scale CRISPR screens identify host factors that promote human coronavirus infection. *Genome Med*. **14**, 10 (2022).

56. Chang, L.-J. & Chen, T.-H. NSP16 2′-O-MTase in coronavirus pathogenesis: Possible prevention and treatments strategies. *Viruses* **13**, 538 (2021).

57. Alshiraihi, I. M., Klein, G. L. & Brown, M. A. Targeting NSP16 methyltransferase for the broad-spectrum clinical management of coronaviruses: managing the next pandemic. *Diseases* **9**, 12 (2021).

58. Li, Q. et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* **182**, 1284–1294.e9 (2020).

59. Syed, A. M. et al. Rapid assessment of SARS-CoV-2 evolved variants using virus-like particles. *Science* **374**, 1626–1632 (2021).

Dae-Kyum Kim[1,2,3,4,5,25], Benjamin Weller [6,25], Chung-Wen Lin[6,25], Dayag Sheykhkarimli [1,2,3,4,25], Jennifer J. Knapp [1,2,3,4,25], Guillaume Dugied [7,8,9,25], Andreas Zanzoni [10], Carles Pons[11], Marie J. Tofaute [12], Sibusiso B. Maseko[13], Kerstin Spirohn [4,14,15], Florent Laval [4,13,14,15,16,17], Luke Lambourne [4,14,15], Nishka Kishore[1,2,3,4], Ashyad Rayhan[1,2,3,4], Mayra Sauer[6], Veronika Young [6], Hridi Halder[6], Nora Marín-de la Rosa[6], Oxana Pogoutse[1,2,3,4], Alexandra Strobel[6], Patrick Schwehn[6], Roujia Li[1,2,3,4], Simin T. Rothballer[6], Melina Altmann[6], Patricia Cassonnet[7,8,9], Atina G. Coté[1,2,3,4], Lena Elorduy Vergara[6], Isaiah Hazelwood[1,2,3,4], Betty B. Liu[1,2,3,4], Maria Nguyen[1,2,3,4], Ramakrishnan Pandiarajan[6], Bushra Dohai[6], Patricia A. Rodriguez Coloma[6], Juline Poirson [1,2,18], Paolo Giuliana [1,2,3,4], Luc Willems[16,17], Mikko Taipale [1,2,13], Yves Jacob [7,8,9], Tong Hao [4,14,15], David E. Hill [4,14,15,26], Christine Brun [10,19,26], Jean-Claude Twizere[4,13,16,26], Daniel Krappmann [12,26], Matthias Heinig[20,21,26], Claudia Falter[6,26], Patrick Aloy[11,22,26], Caroline Demeret[7,8,9,26] ✉, Marc Vidal[4,14,26] ✉, Michael A. Calderwood [4,14,15,26] ✉, Frederick P. Roth [1,2,3,4,23,26] ✉ and Pascal Falter-Braun [6,24,26] ✉

[1]Donnelly Centre for Cellular and Biomolecular Research (CCBR), University of Toronto, Toronto, Ontario, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [3]Lunenfeld-Tanenbaum Research Institute (LTRI), Sinai Health System, Toronto, Ontario, Canada. [4]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, MA, USA. [5]Department of Cancer Genetics and Genomics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA. [6]Institute of Network Biology (INET), Molecular Targets and Therapeutics Center (MTTC), Helmholtz Zentrum München, German Research Center for Environmental Health, Munich-Neuherberg, Germany. [7]Unité de Génétique Moléculaire des Virus à ARN, Département de Virologie, Institut Pasteur, Paris, France. [8]UMR3569, Centre National de la Recherche Scientifique, Paris, France. [9]Université de Paris, Paris, France. [10]Aix-Marseille Université, Inserm, TAGC, Marseille, France. [11]Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute for Science and Technology, Barcelona, Spain. [12]Research Unit Cellular Signal Integration, Institute of Molecular Toxicology and Pharmacology, Molecular Targets and Therapeutics Center (MTTC), Helmholtz Zentrum München, German Research Center for Environmental Health, Munich-Neuherberg, Germany. [13]Laboratory of Viral Interactomes, GIGA Institute, University of Liège, Liège, Belgium. [14]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA. [15]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [16]TERRA Teaching and Research Centre, University of Liège, Gembloux, Belgium. [17]Laboratory of Molecular and Cellular Epigenetics, GIGA Institute, University of Liège, Liège, Belgium. [18]Molecular Architecture of Life Program, Canadian Institute for Advanced Research (CIFAR), Toronto, ON, Canada. [19]CNRS, Marseille, France. [20]Institute of Computational Biology (ICB), Computational Health Center, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich-Neuherberg, Germany. [21]Department of Informatics, Technische Universität München, Munich, Germany. [22]Institució Catalana de Recerca I Estudis Avaçats (ICREA), Barcelona, Spain. [23]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. [24]Microbe-Host Interactions, Faculty of Biology, Ludwig-Maximilians-Universität (LMU) München, Planegg-Martinsried, Germany. [25]These authors contributed equally: Dae-Kyum Kim, Benjamin Weller, Chung-Wen Lin, Dayag Sheykhkarimli, Jennifer J. Knapp, Guillaume Dugied. [26]These authors jointly supervised this work: David E. Hill, Christine Brun, Jean-Claude Twizere, Daniel Krappmann, Matthias Heinig, Claudia Falter, Patrick Aloy, Caroline Demeret, Marc Vidal, Michael A. Calderwood, Frederick P. Roth, Pascal Falter-Braun. ✉e-mail: caroline.demeret@pasteur.fr; marc_vidal@dfci.harvard.edu; michael_calderwood@dfci.harvard.edu; fritz.roth@utoronto.ca; pascal.falter-braun@helmholtz-muenchen.de

Methods

Cloning SARS-CoV-2 ORFs

Two independent SARS-CoV-2 vORF collections were constructed in Gateway entry vectors. The Y2H_{GFP} collection[60] includes all but one (NSP11 was too short for Gateway cloning) codon-optimized ORF of SARS-CoV-2, synthesized based on a published genome[61], which were cloned with and without stop codon, to enable C-terminal fusions. The Y2H_{HIS3} entry clone collection is based on National Center for Biotechnology Information (NCBI) accession number NC_045512.2 and annotation[62]. Y2H_{HIS3} vORFs were synthesized by Twist Bioscience without codon optimization and included 5′ and 3′ linkers with SfiI restriction sites. The 5′ linker incorporates a translational start ATG flanked by BamHI sites; the 3′ linker provides a stop codon flanked by PacI and AsiSI restriction sites. For Y2H_{HIS3}, vORFs were cloned into pENTR223.1 using SfiI restriction cloning, and the alternative ATG was removed by BamHI digest. A total of 28 vORFs were synthesized for Y2H_{GFP} and 27 for Y2H_{HIS3}: NSP1-16 (except NSP11), S, E, M, N and ORFs 3A, 3B, 3D, 6, 7A, 7B, 8, 9B, 9C and 10[62-64] (Supplementary Table 1).

Y2H_{HIS3} vORF entry clones were verified by full-length Sanger sequencing. As NSP10 had a one-base deletion, it was excluded from further experiments. vORFs were moved to the destination vectors pPC86 (N-terminal AD fusion, CEN origin)[3,65] and pHiDEST-DB (N-terminal DB fusion, CEN origin)[4] by Gateway cloning and confirmed by PCR. For Y2H_{GFP}, barcoded 'prey' (pAR068: C-terminal AD fusion, 2μ origin/pHiDEST-AD: N-terminal AD fusion, CEN origin), and 'bait' (pHiDEST-DB: N-terminal DB fusion, CEN origin) destination vectors were generated using published protocols[4], with the integration of the barcode locus at the SacI restriction site as described[26]. Single barcoded plasmid containing colonies were picked, arrayed into 384-well plates with 80 μl LB agar supplemented with 100 μg ml⁻¹ carbenicillin and 35 μg ml⁻¹ chloramphenicol (LB + Carb+CM) per well and incubated at 37°C for 16 h. Barcode sequences were identified using a modified Kiloseq procedure[66] using an Illumina NextSeq 500 and analyzed as previously described[4,26,66]. Y2H_{GFP} vORFs and human ACE2 were moved by Gateway cloning into barcoded destination plasmids[4,26] pHiDEST-AD (N-terminal AD fusion, CEN origin (low copy number)) and pHiDEST-DB (N-terminal DB fusion, CEN origin (low copy number)) such that each ORF was linked to two to six barcodes in every configuration. Gateway cloning was performed individually and for ORF–barcode pairs using Sanger sequencing (TCAG, The Hospital for Sick Children) (Supplementary Table 13).

Generation of HuSCI_{HIS3}

The Y2H_{HIS3} screening pipeline is essentially as previously described[65]. AD-Y and DB-X vORFs were transformed into yeast strains Y8800 (MATa) and Y8930 (MATα), respectively. NSP1 autoactivated as DB fusion and not screened in this orientation. DB-X vORFs were individually mated with 99 pools of ~188 AD-tagged human ORFs each, from human ORFeome v9.1 comprising 17,472 ORFs[26,67] (hORFeome9.1). For the reverse orientation, yeast with 27 AD-Y vORFs were pooled and mated against DB-X hORFeome9.1. Primary screening in both configurations was performed twice to increase sampling sensitivity. Unless otherwise noted, all yeast incubations are at 30°C, overnight without shaking.

For primary screening, saturated haploid AD-Y and DB-X yeast cultures were spotted on top of each other on yeast extract peptone dextrose (YEPD) agar (1%) plates and incubated for 24 h. Yeast were replica plated onto selective synthetic complete media lacking leucine, tryptophan and histidine (SC-Leu-Trp-His) + 1 mM 3-AT (3-amino-1,2,4-triazole)[3,65] (3-AT plates) and incubated for 72 h. From growing spots up to three colonies were picked and cultured in SC-Leu-Trp liquid medium for 2 d. For second phenotyping, cultures were spotted on diploid selection plates, incubated for 2 d and replica plated on 3-AT-plates and SC-Leu-His + 1 mM 3-AT + 1 mg per liter cycloheximide plates to identify spontaneous DB-X autoactivators[2]. Positive scoring colonies (growth

on 3-AT-plates, no growth on cycloheximide plates) were picked, and ORFs were identified by Sanger sequencing[65]. For threefold verification, yeast strains corresponding to the identified human interaction partners were picked from archival glycerol stocks, cultured in liquid medium and mated (as described above) one-by-one against all vORFs, processed as described above and then scored. Colony growth was scored using a custom dilated convolutional neural network[68]. For training, previous datasets of more than 1,500 images of biochemically and functionally validated binary Y2H studies were used[3]. Each image was scaled to achieve equal pixel distance between the yeast spots of different images. The images were cropped and sliced, and the mean grayscale image of all spots on a plate was calculated. With this dataset, a simple front-end prediction module was trained consisting of six dilated convolutional layers with exponential increasing dilation rate and two dense layers at the end. After each layer except the last, a Leaky-ReLU activation was added[69]. The model was optimized with a combination of Softmax and Cross entropy and an Adam Optimizer[70]. The model achieved an accuracy >0.9 during all folds of a tenfold cross-fold validation. All positive scores were confirmed by a trained researcher. The verification step was done in triplicate and protein pairs scoring positive in at least two repeats were considered bona fide Y2H interactors. One representative colony of all interaction pairs was picked from selective plates to confirm the identities of X and Y by Sanger sequencing[65].

Generation of HuSCI_{GFP}

Barcoded ORFeomes. The barcoded human ORFeome consisting of 16,747 fully sequence-verified human ORFs with ~95% ORFs represented by two unique barcodes was previously described[26]. The barcoded bait and prey collections were arranged into a 10-by-10 screening matrix consisting of 10 DB and 10 AD groups, each containing ~1,400 ORFs with two distinct sets of unique barcodes, and ~200 ORFs with a single unique barcode set. Barcoded SARS-CoV-2 plasmids were transformed individually into RY3011 (AD plasmids) and RY3031 (DB plasmids) (genotypes in Supplementary Table 14). Transformed colonies were copied on fresh plates, incubated, scraped off and pooled to make glycerol stocks of all the barcoded SARS-CoV-2 ORFs plus the human ORF ACE2 in each plasmid configuration (with two or more barcodes per ORF).

Mating of pooled haploid yeast. Multiple pooled matings were performed using the frozen haploid pools. Each of the 10 human ORF pools (in C-terminal AD fusion plasmids with 2μ origin; pAR068) were separately mixed with the pool of SARS-CoV-2 ORFs plus human ACE2 (in N-terminal DB fusion plasmids with CEN origins; pHiDEST-DB). A separate mating was done between the SARS-CoV-2 pools in both AD and DB fusion, CEN origin plasmids (pHiDEST-AD, pHiDEST-DB). Negative controls were included in each mating and all matings were calculated to achieve >100× coverage of possible barcode combinations considering viability and mating efficiency. Procedurally, equal amounts of each haploid strain were mixed, the mixture was spread on 2x YEPD plus adenine agar plates (YPAD) and incubated for 24 h. Colonies on each mating plate were collected and re-spread across 20 15 cm SC-Leu-Trp plates supplemented with histidine (8 mM) and incubated for 72 h. These plates were then scraped off to make assay-ready pooled diploid glycerol stocks for each of the 11 groups.

Selection of yeast with interacting pair of DB-X and AD-Y by FACS. Pool of glycerol stocks were inoculated into 1-liter flasks with a starting vCFU of 30 M and incubated at 200 rpm for 24 h. Negative controls were started as 10 ml cultures and processed in parallel. 'Presort' cultures were prepared for each sample (2 × 10 ml cultures with OD₆₀₀ 10) with doxycycline added (10 μg ml⁻¹) to these cultures to induce barcode swapping while these cultures were incubated for 24 h[4]. To prepare for fluorescence-activated cell sorting (FACS), cells were

concentrated by centrifugation ($500 \times g$, 5 min) and resuspended in PBS to a final OD_{600} of 10. Propidium iodide (4 mg liter^{-1}) was added to identify dead yeast cells during FACS. Using the diploid negative control, the FACS gate for GFP-positive cells was set to capture 0.1% of GFP-negative cells, yielding a 0.01% false positive rate. Then, 100 million cells per group were sorted, and GFP-positive cells for each sample were plated on 10 SC-Leu-Trp+Ade+10x His (8 mM) plates and incubated for 72 h. Colonies were collected by scraping, centrifuged and resuspended into 2×10 ml cultures ($OD_{600} = 10$). Doxycycline (10 µg ml^{-1}) was added to induce barcode swapping, and cultures were incubated for 24 h, when plasmid DNA was extracted. Fused barcodes were PCR amplified with primers that attach modified Illumina i5 and i7 adapters to uniquely identify each sample. Following agarose gel analysis of PCR products, the bright band at ~350 bp was purified using a NucleoSpin Gel and PCR Clean-up kit. DNA concentrations were measured for each sample using a Qubit (Invitrogen, Q32851) and, guided by DNA concentration, samples were pooled to ensure equal sequencing depth relative to the number of protein pairs tested. After primer-dimer removal, DNA was quantified by qPCR, and the pooled NGS library was sequenced on an Illumina NextSeq using a mid- or high-output 150-cycles kit.

Read counting based on expected barcodes. The sequencing data were demultiplexed using bcl2fastq2 (v2.20.0.422) provided by Illumina with the following command: 'bcl2fastq -r 10 -p 20 -w 10 –no-lane-splitting –barcode-mismatches 1 –adapter-stringency 0.7 –ignore-missing-bcls –ignore-missing-filter –ignore-missing-positions'. After demultiplexing, the fastq files were aligned to the group specific reference files using bowtie2[71] with the following parameters:

For read 1: -q –norc –local –very-sensitive-local -t -p 23 –reorder.
For read 2: -q –nofw –local –very-sensitive-local -t -p 23 –reorder.

Reference files contained expected barcode sequences for the ORFs in each group. After alignments, reads with mapping quality scores <20 were removed. Following successful BFG barcode recombination[4], paired-end reads map to up-up or dn-dn when an interaction is present. The number of reads mapping to up-up and dn-dn were counted separately and merged as the final read count. The pipeline was implemented in Python v2.7.

Interaction scoring. For virus–host interactions, we used the product of marginal frequencies of bait and prey strains[4] to estimate the abundance of each diploid bait–prey strain in the presort condition ('PreSort'). The interaction score was defined by

$$IS_{ij} = \frac{f_{ij}^{GFP}}{f_{ij}^{PreSort}}$$

$$f_{ij}^{PreSort} = \sum_i c_{ij}^{PreSort} / \sum_j \left[\sum_i c_{ij}^{PreSort} \right]$$

$$f_{j}^{PreSort} = \sum_j c_{ij}^{PreSort} / \sum_j \left[\sum_j c_{ij}^{PreSort} \right]$$

$$f_{j}^{PreSort} = \max \left(f_i^{PreSort}, f_{AD}^{Floor} \right) \times \max \left(f_j^{PreSort}, f_{DB}^{Floor} \right)$$

$$f_{AD}^{Floor} = 10^{-5} f_{DB}^{Floor} = 10^{-4}$$

$$f_{ij}^{GFP} = c_{ij}^{GFP} / \sum_{ij} c_{ij}^{GFP}$$

with the following variables: c, read count; i, AD barcode count; j, DB barcode count; f, frequency.

For every DB barcode, we used the 960 AD null barcodes to define the thresholds leading to a 1% false positive rate. An interaction was accepted as positive only if the ORF pair interaction score was above this threshold for two or more barcode pairs. For intraviral screening, we accepted as interactions those protein pairs for which the frequency of barcode pairs was 1,000 times greater than the median frequency of

the corresponding DB barcode for three or more independent barcode pairs, similar to the scoring method previously used for BFG-Y2H with HIS3-based growth selection[4].

Pairwise retesting. Candidate interaction pairs for HuSCI$_{GFP}$ were verified in a pairwise HIS3 growth-based Y2H assay as described above (Y2H$_{HIS3}$ verification step), with minor modifications. Barcode replicates of candidate human AD-Y and viral DB-X were pooled prior to mating. vORFs NSP1 and NSP12 were omitted from this retesting due to DB autoactivation. After mating, colonies were replica plated on SC-Leu-Trp-His and 3AT-plates. After 72–96 h of yeast growth, these pairwise tests were scored according to the standardized scoring method used for the Y2H$_{HIS3}$ screen[3,65]. Interaction pairs scoring ≥ 3 were considered bona fide Y2H interactions.

Estimating completeness using the interactome framework

Assay sensitivity (S_a) is defined as the fraction of true interactions that can be detected by a given assay. Sampling sensitivity (S_s) is defined as a fraction of detectable true interactions that can be recovered by the pipeline used. Overall sensitivity of a given screen S can be calculated as $S = S_a \times S_s$. In pairwise settings $S_s = 1$ and the assay sensitivity is given by the fraction of hsPRS-v1/v2 pairs that score positive. Y2H$_{HIS3}$ was benchmarked previously[5] and has an assay sensitivity of $S_{a\text{-}HIS3} = 21.7\%$. Sampling sensitivity of Y2H$_{HIS3}$ after two repeats in two orientations has been shown to be $S_{s\text{-}HIS3} = \sim60\%$[65], yielding a screening sensitivity of $S_{HIS3} = S_{a\text{-}HIS3} \times S_{s\text{-}HIS3} = 0.217 \times 0.6 = 13\%$. Given that Y2H$_{HIS3}$ screen had a search space completeness of 83% ($T_{HIS3} = 83\%$), the overall completion of HuSCI$_{HIS3}$ is $C_{HIS3} = T_{HIS3} \times S_{HIS3} = 0.83 \times 0.13 = 10.8\%$.

A different version of Y2H$_{GFP}$ using low-copy plasmids and N-terminally fused hybrid proteins (lcnY2H$_{GFP}$) was benchmarked using 84 pairs of hsPRS-v1 and 92 pairs of hsRRS-v1. Flow cytometry was used to score interactions based on percentage of singlets in GFP-positive gate, which was set using empty bait and prey constructs. In addition, lcnY2H$_{GFP}$ was benchmarked in a pooled setting using all possible combinations of proteins constituting 78 hsPRS-v2 and 77 hsRRS-v2 pairs supplemented with a set of 14 pairs of Y2H-positive controls defined as calibration set[4]. The experiment was carried out and interactions were scored as described above, except that no empirical null distribution was used. lcnY2H$_{GFP}$ recovered 12 out of 82 ($S_{a\text{-}lcnGFP} = 15\%$) hsPRS-v1 pairs when tested in a pairwise single bait–prey configuration and 8 of 92 (9%, $S_{s\text{-}lcnGFP} = 9/15 = 60\%$) hsPRS-v2 + calibration set pairs when tested in a pooled single bait–prey configuration, yielding $S_{lcnGFP} = S_{a\text{-}lcnGFP} \times S_{s\text{-}lcnGFP} = 0.15 \times 0.6 = 9\%$. It has been previously shown that using high-copy C-terminal fusions increases sensitivity by ~33% without affecting precision[26]. Thus, screening sensitivity of Y2H$_{GFP}$ was modeled from that of lcnY2H$_{GFP}$ as $S_{GFP} = S_{lcnGFP} \times 1.33 = 9\% \times 1.33 = 12\%$. Given that Y2H$_{GFP}$ covered 70% ($T_{GFP} = 70\%$) of all possible virus–human protein combinations, the completion level of the Y2H$_{GFP}$ dataset is $C_{GFP} = T_{GFP} \times S_{GFP} = 0.70 \times 0.12 = 8.4\%$. Only 4 out of 28 (14.2%) hsPRS-v1 pairs detected by the union of Y2H$_{HIS3}$ and lcnY2H$_{GFP}$ were detected with both methods, indicating a high degree of orthogonality (that is, different detection profiles of the methods used). In addition, Y2H$_{GFP}$ implemented in this study includes further differences such as high-copy and C-terminal fusion constructs for human proteins. Therefore, we conservatively estimate 90% orthogonality between Y2H$_{HIS3}$ and Y2H$_{GFP}$ (that is, ~90% of detected interactions are different: $O_{HIS3+GFP} = 90\%$). Thus, we estimate that the fraction of all true interactions captured by our merged interactome maps is $C_{HIS3+GFP} = (C_{HIS3} + C_{GFP}) \times O_{HIS3+GFP} \cong (0.108 + 0.084) \times 0.9 = 17.3\%$. Given the uncertainties associated with derivation of screening sensitivity, we estimate lower and higher bounds to be 15% ($S_{GFP} = 9\%$, excluding inferred gain in sensitivity due to high-copy C-terminal fusions) and 22% ($S_{GFP} = 13.5\%$, $S_{s\text{-}HIS3} = 70\%$ and $O_{HIS3+GFP} = 100\%$), respectively.

Pairwise Y2H testing of previously identified SARS-CoV-1 interactions

We identified 97 unique curated binary interactions with SARS-CoV-1 and human interaction partners[8] (Supplementary Table 2). For 77 of these, reagents to test interactions with SARS-CoV-2 orthologues were available in the barcoded human ORFeome. These involved 63 human proteins, 60 of which were covered by two barcode sets and three by a single barcode set. These were tested according to the 'pairwise retesting' protocol (above). Successful interactions were indicated by colony growth of both replicates in either condition.

Pairwise Y2H testing with SARS-CoV-2 variants

Lineage-defining mutations for the SARS-CoV-2 'variants of concern' as defined by the Centers for Disease Control and Prevention (Alpha, Beta, Gamma and Delta) were obtained from CoV-Spectrum[72,73] and mapped to the SARS-CoV-2 reference genome (NCBI accession number NC_045512.2). To generate variant ORFs, Y2H$_{HIS3}$ plasmids were used as template for mutation PCR (primers in Supplementary Table 12). Mutation PCR reaction products were transformed and sequence verified. Plasmids containing the desired mutation were directly transformed into yeast and processed in pairwise mating as described above. A complete list of mutations generated is shown in Supplementary Table 12. SARS-CoV-2 proteins for which interactions were identified in AD-fusions (N and E) were tested only against the identified interactors. All other variant proteins were tested against all HuSCI interactors. In total, 19 individual mutations in 14 unique variant proteins from 9 different viral proteins were tested. Four proteins with 8 cloned variants had interactors in HuSCI$_{HIS3}$, 1 protein with a single cloned variant had interactors in HuSCI$_{GFP}$ and 4 proteins with 5 variants had no HuSCI interactors.

yN2H validation

Using Gateway cloning, ORFs from the indicated subsets (Supplementary Table 3) were transferred into pDEST-N2H plasmids (pDEST-N2H-N1, -N2, -C1, and -C2) containing a *LEU2* (N1/C1 vectors) or a *TRP1* (N2/C2 vectors) auxotrophy marker and transformed into haploid *Saccharomyces cerevisiae* Y8800 (MATa) and Y8930 (MATα) strains. For cross-plate calibration, two protein pairs from the hsPRS-v2, with different N2H signal intensities, were included in duplicate on every plate (NCBP1/NCBP2 and SKP1/SKP2). Virus–human protein pairs were randomly distributed across the plates and tested together with hsPRS-v2/hsRRS-v2, which were in separate plates.

Overnight-grown haploid cultures were mated by mixing 5 µl of each haploid strain in 160 µl YEPD medium followed by overnight incubation. To measure background, all interactor ORFs were also mated with yeast with empty F1 or F2 plasmids. After mating, 10 µl culture each was inoculated into 160 µl SC-Leu-Trp and grown overnight, and then 50 µl was reinoculated into 1.2 ml SC-Leu-Trp and incubated for 24 h while shaking at 900 rpm. Cells were harvested (6,000 x *g*, 15 min), and the supernatant was discarded. Each yeast cell pellet was fully resuspended in 100 µl NanoLuc Assay solution[6]. Homogenized solutions were transferred into white flat-bottom 96-well plates and incubated in the dark (for 1 h at room temperature). Luminescence was evaluated for each sample with 2 s integration time. To score X–Y protein pairs, a normalized luminescence ratio (NLR) was calculated corresponding to the raw luminescence value of the tested pair (X-Y) divided by the maximum luminescence value from one of the two controls (X-Fragment 2 or Fragment 1-Y)[6]. The 1% RRS threshold was based on the vhRRS and determined using the R quantile function.

Enrichment of previously known, phospho-regulated or RNA-binding host targets

From IntAct[8] (version: April 28, 2020), 2,151 human proteins reported to have binary interactions with any virus protein were defined as 'previously known host targets'. 2,005 of these ORFs were interrogated by our experiment, and further considered. HuSCI contained 61 previously known host targets. 2,254 human proteins that change phosphorylation changes upon SARS-CoV-2 infection were identified from A549 and Vero E6 cell lines[9,10], of which 2,007 were interrogated by our experiment and 37 are in HuSCI. 139 experimentally identified human proteins specifically bound to SARS-CoV-2 RNA (vRICs) and 335 human proteins with altered RNA-binding activity upon SARS-CoV-2 infection (cRICs) were obtained from a recent RNA-interactome study[11]. Then, 121 vRICs and 294 cRICs were interrogated by our experiment; 5 HuSCI proteins were vRICs, and 13 HuSCI proteins were cRICs. All the observations were tested for enrichment using Fisher's exact tests and by permutation tests with 10,000 permutations.

GO enrichment analysis

gProfiler[74] (database versions: Ensembl 104, Ensembl Genomes 51 and Wormbase ParaSite 15) was applied to identify enriched functional categories in HuSCI, AP-MS[9,12–15] and BioID studies[16–18]. The hORFeome9.1, which was used for contactome mapping, served as the background for HuSCI, otherwise the universal annotated human genes. 'Inferred from electronic annotations' annotations were excluded. Adjusted *P* values were calculated using the Benjamini–Hochberg procedure. Functional terms with a hypergeometric *P* < 0.05 and term size between 5 and 1,000 were collected and enrichment calculated as the ratio between observed and expected gene counts. To categorize HuSCI host proteins, five meta categories inspired by the functional enrichment analysis results were used, namely 'immune response' (GO:0006955), 'viral process' (GO:0016032), 'protein ubiquitination' (GO:0016567), 'cytoskeleton' (GO:0005856) and 'vesicle-mediated transport' (GO:0016192). Human proteins related to these categories were obtained from the AmiGO 2 (ref. [75]) (July 2021), and HuSCI host proteins were categorized based on their annotation to these meta categories.

Domain enrichment of host interacting proteins

Structural domains in human targets were identified from Pfam release 34.0 (ref. [76]) (March 2021). Interactions of viral proteins with human interactors that have common domains were defined as shared-domain interactions and counted for HuSCI. The procedure was repeated for 1,000 randomized HuSCI networks (degree-preserved random rewiring). The significance of every viral protein–human domain was assessed by Fisher's exact tests (Supplementary Table 6) using the number of V-D, V-!D, !V-D, and !V-D interacting pairs, in which V and D correspond to the viral protein and human domain of interest, and !V and !D to the rest of viral proteins and domains in the HuSCI network, respectively. We identified as enriched associations those with at least two V-D interactions and *P* < 0.05. We repeated the process for 1,000 randomized HuSCI networks (see above). Multiple domain copies in a given human protein were counted once.

NF-κB reporter assays

HEK293 (RRID: CVCL_0045, DSMZ) were cultured in complete DMEM (high glucose) supplemented with 10% fetal calf serum, 100 U ml^{-1} penicillin and 100 µg ml^{-1} streptomycin and maintained in humidified atmosphere at 5% CO_2 at 37°C. For the reporter assay, 1×10^6 HEK293 cells were seeded in a 60-mm cell culture dish one day before transfection. Transfection was done using the calcium phosphate protocol using 10 ng NF-κB reporter plasmid (6 × NF-κB firefly luciferase pGL2), 50 ng pTK reporter (Renilla luciferase) and expression vectors (Flag-IKKb (pRK5), Flag-A20 (pEF4) and SARS-CoV-2 constructs (pMH)) using a total of up to 6 µg DNA. Briefly, the DNA was diluted in 200 µl 250 mM $CaCl_2$ solution (Carl Roth, 5239.1), vortexed and added dropwise to 200 µl 2 × HBS (50 mM HEPES (pH 7.0), 280 mM NaCl, 1.5 mM $Na_2HPO_4 \times 2H_2O$, pH 6.93) while gently vortexing. After 15-min incubation at room temperature, the mix was added dropwise to cell culture dishes. Transfection media was replaced after 6-h incubation with complete DMEM. Then, 24 h after transfection cells were stimulated with 20 ng ml−1 TNF-α for 4 hours. Luciferase activity was measured

using the dual luciferase reporter kit (Promega, E1980) according to the manufacturer's protocol. The firefly and Renilla luminescence was determined with a luminometer (Berthold Centro LB960 microplate reader, software MikroWin 2010) and quantified in relative light units (RLU). NF-κB induction was specified as the ratio of firefly luminescence (RLU) to Renilla luminescence (RLU). Significance of relative NF-κB transcriptional activity was assessed via one-way ANOVA with Dunnett's multiple comparisons. Data evaluation was performed in GraphPad Prism v7.04.

Protein expression was verified by western blot of lysates. Briefly, proteins were separated by SDS-PAGE and transferred on polyvinylidene fluoride membranes. Membranes were blocked with 5% milk in 1× PBS + 0.1% Tween-20 (PBS-T) for 1 h at room temperature. Primary antibodies in 2.5% milk in PBS-T were incubated overnight at 4 °C, the membranes were washed three times with PBS-T and secondary antibodies were incubated (1.25% milk/PBS-T) for 1 h at room temperature. Anti-actin beta (SCBT, sc-47778), anti-FLAG M2 (Sigma-Aldrich, F3165) and anti-HA (Sigma-Aldrich, 11583816001, RRID:AB_514505) were used at a 1:1,000 dilution. Secondary antibody (Jackson ImmunoResearch, Jim-715-035-150) was used at a 1:10,000 dilution. For detection of horseradish peroxidase-catalyzed enhanced chemiluminescence, LumiGlo reagent (CST, 7003S) was used.

For generation of *IKBKG* KO HEK293 cells, oligonucleotides coding sgRNAs targeting exon 3 (5′-TGCATTTCCAAGCCAGCCAG-3′) or exon 2 (5′- GCTGCACCATCTCACACAGT-3′) were cloned into px458 (Addgene, 48138). HEK293 were transfected with 5 μg plasmid by standard calcium phosphate transfection. After one day, GFP-positive cells were sorted with a MoFlo cell sorter (Beckman Coulter, Cytomation) and seeded in 96-well plates at dilutions of 0.5–5.0 cells per well. Single-cell clones were expanded and screened for loss of IKBKG expression by western blot (RRID: AB_2124846). IKBKG-negative clones were verified by amplifying and sequencing a region of genomic DNA encompassing the sites targeted by PCR (exon 3: forward primer 5′-CTGGCCAACACGTACTTTTA-3′, reverse primer 5′-GGTTACGGTGAGCGAAGGCTC-3′; exon 2: forward primer 5′- CTGACATCTCCCTCCACAAAC-3′ and reverse primer 5′-GGAGCTGGAATGAACCTTCC-3′).

Functional effects on viral replication
Selection of host-target candidates. To evaluate if identified host targets are involved in viral replication, the following HuSCI proteins involved in host immune regulation[77] and viral life cycle regulation[51,78–80] by enriched GO terms in this study were selected: G3BP1, G3BP2, TRAF2, USP25, EIF2AK2, REL, IKBKG and KLC1.

Engineering of hACE2-expressing cells. A549 cells were seeded at 5 × 10⁵ cells per well in six-well cell culture plates and cultured in DMEM with 10% FCS and 1% penicillin/streptomycin at 37 °C and 5% CO₂ (standard media). After 24 h culture medium was replaced by fresh medium containing 4.5 × 10⁷ transduction units hACE2 lentivirus per well and incubated for 4 hours at 37 °C and 5% CO₂. The lentiviral inoculum was then replaced with 2 ml DMEM 10% FCS and 1% penicillin/streptomycin. After 24 h, the transduction was repeated with the same steps as above. Cell surface expression of hACE2 was monitored by FACS using the AttuneNxT Flow Cytometer (Thermo Fisher Scientific) and results were analyzed with FlowJo v10 Software (BD Life Sciences). The resulting cells are referred to as A549-hACE2.

Generation of KO cell lines. KO cells were generated using the target-specific CRISPR-Cas9-HDR (homology-directed recombination) KO directed technology developed by Santa Cruz Biotechnology, which enables selection of KO cells with puromycin and red fluorescent protein (Supplementary Table 15). Briefly, A549-hACE2 cells were seeded at 2.5 × 10⁶ cells in T25 flasks and standard media. After 24 h, cells were cotransfected with 7.5 μg each of KO and HDR plasmids for

the previously described targets and 15 μg KO plasmid for the mock KO, from Santa Cruz Biotechnology using FuGene (Promega, E2312). After 72 h, KO cells were selected with 2 μg/ml puromycin (Invivogen, ant-pr-1) for 3 d, and mock KO cells were treated with the same volume of Hepes solution (Sigma-Aldrich, 51558). One week later, red fluorescent protein-positive cells were sorted by flow cytometry. DNA from 2 × 10⁶ cells was extracted and region of interest was amplified for each KO, except KLC1, in a 25-μl PCR using 50 ng genomic DNA and using one primer in the genomic DNA and one primer in the insert (primers are listed in Supplementary Table 15). KLC1 KO was verified by amplifying the sg-directed Cas9 region that had no corresponding HDR with one primer on each side of the region; the PCR product was purified using Nucleospin Gel and PCR Clean-up (Machery-Nagel, 11992242) and KO confirmed by Sanger sequencing.

Assessment of SARS-CoV-2 infection in A549-hACE2 KO versus wild-type cells. Wild-type and KO A549-hACE2 cells were seeded at 1 × 10⁶ cells per well in 12-well plates and standard media. After 24 h, cells were infected at a multiplicity of infection (MOI) of 10⁻³, with SARS-CoV-2 isolate hCoV19/France/GE1973/2020 (*n* = 3, biological replicates). Total RNA was extracted from infected cells at 72 h after infection, and SARS-CoV-2 replication was assessed by RT-qPCR using Orf1ab primers (5′-ATGAGCTTAGTCCTGTTG-3′; 3′-CTCCCTTTGTTGTGTTGT-5′) (*n* = 9, three technical replicates per biological replicate). GAPDH was used for normalization. Viral RNA was quantified according to the ΔΔCt standard method[81]. The effect of gene KO on viral replication was determined using the wild-type ORF1ab RNA level as a control as shown in the following equation: $2^{-(\Delta\Delta Ct)} = 2^{-(\Delta Ct_{KO} - \Delta Ct_{WT})}$. Significance of the KO effect was calculated against the mock KO using an ordinary one-way nonparametric ANOVA Kruskal-Wallis with Dunn's multiple comparisons test using GraphPad Prism v9.

Assessment of the viability of the KO cell lines. A total of 8.0 × 10⁵ cells of each KO cell line were seeded in a white 96-well plate and incubated at 37 °C and 5% CO₂ for 24 h. Cell media was replaced with DMEM and incubated at 37 °C and 5% CO₂ for 72 h. Cell viability was measured using Cell Titer-Glo Luminescent Cell Viability Assay kit (Promega, G7750). Luminescence was measured on a Centro XS luminometer (Berthold; integration time, 0.5 s). Wild-type cells served as the reference and significance of cell viability was calculated against the mock KO using an ordinary one-way nonparametric ANOVA Kruskal–Wallis with Dunn's multiple comparisons test using GraphPad Prism v9.

Genes ranked by number of publications
Publication counts are derived from the gene2pubmed file from NCBI, downloaded on 16 November 2021. Only protein-coding genes were considered. For visualization, but not statistical assessment, of genes with equal numbers of publications, order was determined by random shuffling. *P* values were calculated by Mann–Whitney *U* test, with Bonferroni correction. Black dots indicate the mean; error bars represent the 95% confidence interval generated from 1,000 bootstrap samples.

Tissue specificity analysis
The Tissue Atlas dataset was obtained from the HPA database[21] (version 2021.04.09). The HPA categories 'tissue enriched', 'group enriched' and 'tissue enhanced' were combined with 'tissue-specific', 'low tissue specificity' was denoted as 'common' and the 'not detected' category was not included in this analysis. A total of 11,069 of 19,670 genes (56.3%) in the HPA dataset were defined as tissue specific, and 8,385 of 19,670 genes (42.6%) showed common expression profiles. Tissue distribution differences were determined using Fisher's exact test with Bonferroni correction.

SARS-CoV-2 organotropism data were obtained from post mortem examinations[22,82]. The RNA tissue-specific NX value (normalized transcripts per million) was extracted and used to denote whether the gene is specifically expressed in a given tissue. Tissues from the

Tissue Atlas were combined into organ systems and used to assess host-target tissues. Significance was evaluated by Fisher's exact test with Bonferroni correction.

Identification of genetic variation in host targets and network communities

Host network communities were identified using the OCG hierarchical community clustering algorithm on the Human Reference Interactome[26,83] as implemented in the linkcomm R package (V1.0-13) using 'centered cliques' as initial class system[84]. A total of 3,603 communities with a minimum size of 4 were found, of which 204 contained a significant number of virus interactors (that is, were significantly targeted) (nominal $P < 0.05$, Fisher's exact test; Supplementary Table 8). A community was annotated to a function if a GO term was enriched (FDR < 0.05) or if ≥20% or ≥30% of the annotated constituent proteins shared an annotation[85] (Supplementary Table 8). From AP-MS-based association studies[9,12–15], 57, 43, 18 and 17 significantly targeted communities were found, respectively (nominal $P < 0.05$, Fisher's exact test; Supplementary Table 8).

Uniformly processed GWAS summary statistics were downloaded for 114 traits from the GTEx GWAS analysis[41,86]. MAGMA[87] analysis was implemented in R 3.6.1 and consists of three steps: first, GWAS summary statistics across all single-nucleotide polymorphisms (SNPs) within a gene region are aggregated into a gene-level association P value. Next, the gene-level P value is transformed to a z-score (using the inverse normal cumulative distribution function). Finally, z-scores across all genes are modeled as a function of gene set membership and the default gene-level covariates (gene size in number of SNPs, the gene density (a measure of within-gene linkage disequilibrium), the inverse mean minor allele count) using a linear model. Association between gene set membership and GWAS z-scores is tested based on the null hypothesis beta = 0 for the coefficient associated with the gene set membership indicator variable. All targets, and the targeted network communities, were considered gene sets. Entrez gene IDs were used on the human genome assembly 38. Individual MAGMA analyses were performed for each trait based on summary statistics and linkage disequilibrium structure from the 1,000 genomes European reference panel always conditioning on default gene-level covariates (for example, gene length). For each gene set, standard error normalized beta coefficients constituted the association score, with larger values indicating greater chance of getting significant association. Following Benjamini–Hochberg multiple hypothesis correction, gene set–trait associations with FDR < 0.05 were selected. These pairs were subjected to follow-up analysis. SNPs localizing within genes of enriched gene sets were selected, and genes containing SNPs with GWAS $P < 5.0 \times 10^{-8}$ were selected for the enriched traits, which were considered 'GWAS hits'. As control the analysis was repeated for the 3,399 network communities that were not significantly targeted (Supplementary Table 8). For both targeted and non-targeted communities the probability of observing traits that are linked to COVID-19 outcomes was assessed. A literature survey identified 35 traits clinically linked to COVID-19 (score 2 in Supplementary Table 8), 18 'related to immune function' and 61 without connection. For the enrichment analysis we focused on the 'COVID-linked' traits; traits 'related to immune function' are also indicated in Fig. 3. Finally, Fisher's exact test was used to assess the significance traits being linked to COVID-19 (score 2) vs not (scores 0 and 1) in traits that are associated with not-virus- targeted communities ($P = 0.5$) vs virally targeted communities ($P = 0.01$). For the control analysis of AP-MS targeted communities, only genetic variation related to COVID-19 severity was evaluated. The contactome-targeted communities with significant GWAS trait associations were numbered 1–31.

Small-molecule inhibition

Remdesivir (Bio-Techne, 7226/10) and USP25/28 inhibitor AZ1 (Bio-Connect, HY-117370-5mg) were dissolved in DMSO. HEK293-ACE2 and Vero E6 (3×10^4 cells per well) were plated in white 96-well plates. After 24 h, cells were infected with SARS-CoV-2 (ref. [54]) (0.01 MOI) containing a nanoluciferase reporter and treated with the compounds in a 12-point twofold dilution series with 0–10 μM concentration. Each condition was done in triplicate, except for AZ1, which was done in quadruplicate for HEK293-ACE2 and one replicate for Vero E6. Cells were cultured for 24 h, and luminescence was quantified[88]. Cell viability was measured using the Cell Titer-Glo Luminescent Cell Viability Assay kit (Promega, G7750). EC_{50} values were calculated via the variable slope model in GraphPad Prism v9.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The protein-protein interaction (PPI) data from this publication have been submitted to the IMEx (http://www.imexconsortium.org) consortium through IntAct and assigned the identifier IM-28880 (ref. [89]). All data from the study are included in the article and associated files. Source data are provided with this paper.

The following data were obtained from the respective original publications: phosphorylation changes upon SARS-CoV-2 infection[9,10]; RNA-binding changes upon SARS-CoV-2 infection[11]; AP-MS virus–host association data: Gordon et al.[12,13], Stukalov et al.[9], Li et al.[14], Nabeel-Shah et al.[15]; BioID virus–host proximity data: Laurent et al.[16], St-Germain et al.[17] Samavarchi-Tehrani et al.[18]; human expression data: Human Proteome Atlas[21], SARS-CoV-2 organotropism[22,82]; human host interactome: HuRI[26]; GWAS data for severe COVID-19 illness[32,33]; and GWAS summary statistics for 114 traits: doi:10.5281/ZENODO.3518299. Interaction data for other viruses were downloaded from IntAct[8] (version: 28 April 2020). Publication counts were downloaded from gene2pubmed (NCBI) on 16 November 2021. Source data are provided with this paper.

Code availability

All source code related to this paper is available via GitHub (https://github.com/INET-HMGU/SARS-CoV-2-contactome)[90].

References

60. Kim, D.-K. et al. A comprehensive, flexible collection of SARS-CoV-2 coding regions. *G3* **10**, 3399–3402 (2020).
61. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
62. Wu, A. et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* **27**, 325–328 (2020).
63. Jungreis, I. et al. Conflicting and ambiguous names of overlapping ORFs in the SARS-CoV-2 genome: A homology-based resolution. *Virology* **558**, 145–151 (2021).
64. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
65. Altmann, M., Altmann, S., Falter, C. & Falter-Braun, P. High-quality yeast-2-hybrid interaction network mapping. *Curr. Protoc. Plant Biol.* **3**, e20067 (2018).
66. Weile, J. et al. A framework for exhaustively mapping functional missense variants. *Mol. Syst. Biol.* **13**, 957 (2017).
67. The ORFeome Collaboration. The ORFeome collaboration: a genome-scale human ORF-clone resource. *Nat. Methods* **13**, 191–192 (2016).
68. Fisher, Y. & Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. CoRR abs/1511.07122 (JMLR.org, 2016): n. pag.
69. Maas, A. L, Hannun, A. Y & Ng, A. Y. Rectifier nonlinearities improve neural network acoustic models. *Proceedings of the 30th International Conference on Machine Learning*, **30** (Atlanta, GA, 2013).

70. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at *arXiv* https://arxiv.org/abs/1412.6980 (2014).

71. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

72. Chen, C. et al. CoV-Spectrum: analysis of globally shared SARS-CoV-2 data to identify and characterize new variants.*Bioinformatics* **38**, 1735–1737 (2021).

73. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).

74. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res* **35**, W193 (2007).

75. Carbon, S. et al. AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288 (2009).

76. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res*. **49**, D412 (2021).

77. Shin, C. et al. MKRN2 is a novel ubiquitin E3 ligase for the p65 subunit of NF-κB and negatively regulates inflammatory responses. *Sci. Rep.* **7**, 46097 (2017).

78. Götte, B. et al. Separate domains of G3BP promote efficient clustering of alphavirus replication complexes and recruitment of the translation initiation machinery. *PLoS Pathog.* **15**, e1007842 (2019).

79. Hosmillo, M. et al. Noroviruses subvert the core stress granule component G3BP1 to promote viral VPg-dependent translation. *eLife* **8**, e46681 (2019).

80. Liu, S., Dominska-Ngowe, M. & Dykxhoorn, D. M. Target silencing of components of the conserved oligomeric Golgi complex impairs HIV-1 replication. *Virus Res.* **192**, 92–102 (2014).

81. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402–408 (2001).

82. Meinhardt, J. et al. Olfactory transmucosal SARS-CoV-2 invasion as a port of central nervous system entry in individuals with COVID-19. *Nat. Neurosci.* **24**, 168–175 (2020).

83. Becker, E., Robisson, B., Chapple, C. E., Guénoche, A. & Brun, C. Multifunctional proteins revealed by overlapping clustering in protein interaction network. *Bioinformatics* **28**, 84–90 (2012).

84. Kalinka, A. T. & Tomancak, P. linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. *Bioinformatics* **27**, 2011–2012 (2011).

85. Chapple, C. E. et al. Extreme multifunctional proteins identified from a human protein interaction network. *Nat. Commun.* **6**, 7412 (2015).

86. Barbeira, A. N. *et al*. GWAS and GTEx QTL integration. *Zenodo* https://doi.org/10.5281/ZENODO.3518299 (2019).

87. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

88. Coutant, E. P. et al. Bioluminescence profiling of NanoKAZ/NanoLuc luciferase using a chemical library of coelenterazine analogues. *Chemistry* **26**, 948–958 (2020).

89. Kim, D.K. et al. IM-28880. IMEx. https://www.ebi.ac.uk/legacy-intact/query/pubid:unassigned2933;jsessionid=E9D9D501AAC618B88078DBD0BD47AEFA?conversationContext=1 (2022).

90. Kim, D.K. et al. SARS-CoV-2-contactome. GitHub. https://github.com/INET-HMGU/SARS-CoV-2-contactome (2022).

91. Barron, E. et al. Associations of type 1 and type 2 diabetes with COVID-19-related mortality in England: a whole-population study. *Lancet Diabetes Endocrinol.* **8**, 813–822 (2020).

92. Leong, A. et al. Cardiometabolic risk factors for COVID-19 susceptibility and severity: a Mendelian randomization analysis. *PLoS Med.* **18**, e1003553 (2021).

93. Nikniaz, Z., Somi, M. H., Dinevari, M. F., Taghizadieh, A. & Mokhtari, L. Diabesity associates with poor COVID-19 outcomes among hospitalized patients. *J. Obes. Metab. Syndr.* **30**, 149–154 (2021).

94. Aung, N., Khanji, M. Y., Munroe, P. B. & Petersen, S. E. Causal inference for genetic obesity, cardiometabolic profile and COVID-19 susceptibility: a Mendelian randomization study. *Front. Genet.* **11**, 586308 (2020).

95. Freuer, D., Linseisen, J. & Meisinger, C. Impact of body composition on COVID-19 susceptibility and severity: a two-sample multivariable Mendelian randomization study. *Metabolism* **118**, 154732 (2021).

96. Wang, C. et al. Red cell distribution width (RDW): a prognostic indicator of severe COVID-19. *Ann. Transl. Med.* **8**, 1230 (2020).

97. Ouyang, S.-M. et al. Temporal changes in laboratory markers of survivors and non-survivors of adult inpatients with COVID-19. *BMC Infect. Dis.* **20**, 952 (2020).

98. Kearns, S. M. et al. Reduced adiponectin levels in patients with COVID-19 acute respiratory failure: a case-control study. *Physiol Rep.* **9**, e14843 (2021).

99. Hypothyroidism is associated with prolonged COVID-19-induced anosmia: a case-control study. *J. Neurol. Neurosurg. Psychiatry* **20**, jnnp–2021–326587 (2021).

100. Brancatella, A. et al. Subacute thyroiditis after SARS-CoV-2 infection. *J. Clin. Endocrinol. Metab.* **105**, dgaa276 (2020).

101. Nemani, K. et al. Association of psychiatric disorders with mortality among patients with COVID-19. *JAMA Psychiatry* **78**, 380–386 (2021).

102. Zhu, Z. et al. Association of obesity and its genetic predisposition with the risk of severe COVID-19: analysis of population-based cohort data. *Metabolism* **112**, 154345 (2020).

103. Derikx, L. A. A. P. et al. Clinical outcomes of COVID-19 in patients with inflammatory bowel disease: a nationwide cohort study. *J. Crohns. Colitis* **15**, 529–539 (2021).

104. Dar, H. Y., Azam, Z., Anupam, R., Mondal, R. K. & Srivastava, R. K. Osteoimmunology: the between bone and immune system. *Front. Biosci.* **23**, 464–492 (2018).

Acknowledgements

Author contributions

Conceptualization: D.-K.K., B.W., C.-W.L., D.S., J.J.K., C.F., F.P.R. and P.F.-B.; SARS-CoV-2 clone construction: D.-K.K., B.W., J.J.K., M.S., H.H., N.M.R., O.P. and P.C.; yeast two-hybrid screening: D.-K.K., B.W., D.S.,

Funding

Competing interests

F.R. and M.V. are advisors and shareholders of SeqWell, Inc. (Beverly, MA, USA).

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41587-022-01475-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-022-01475-z.

Correspondence and requests for materials should be addressed to Caroline Demeret, Marc Vidal, Michael A. Calderwood, Frederick P. Roth or Pascal Falter-Braun.

Peer review information *Nature Biotechnology* thanks Ulrich Stelzl and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
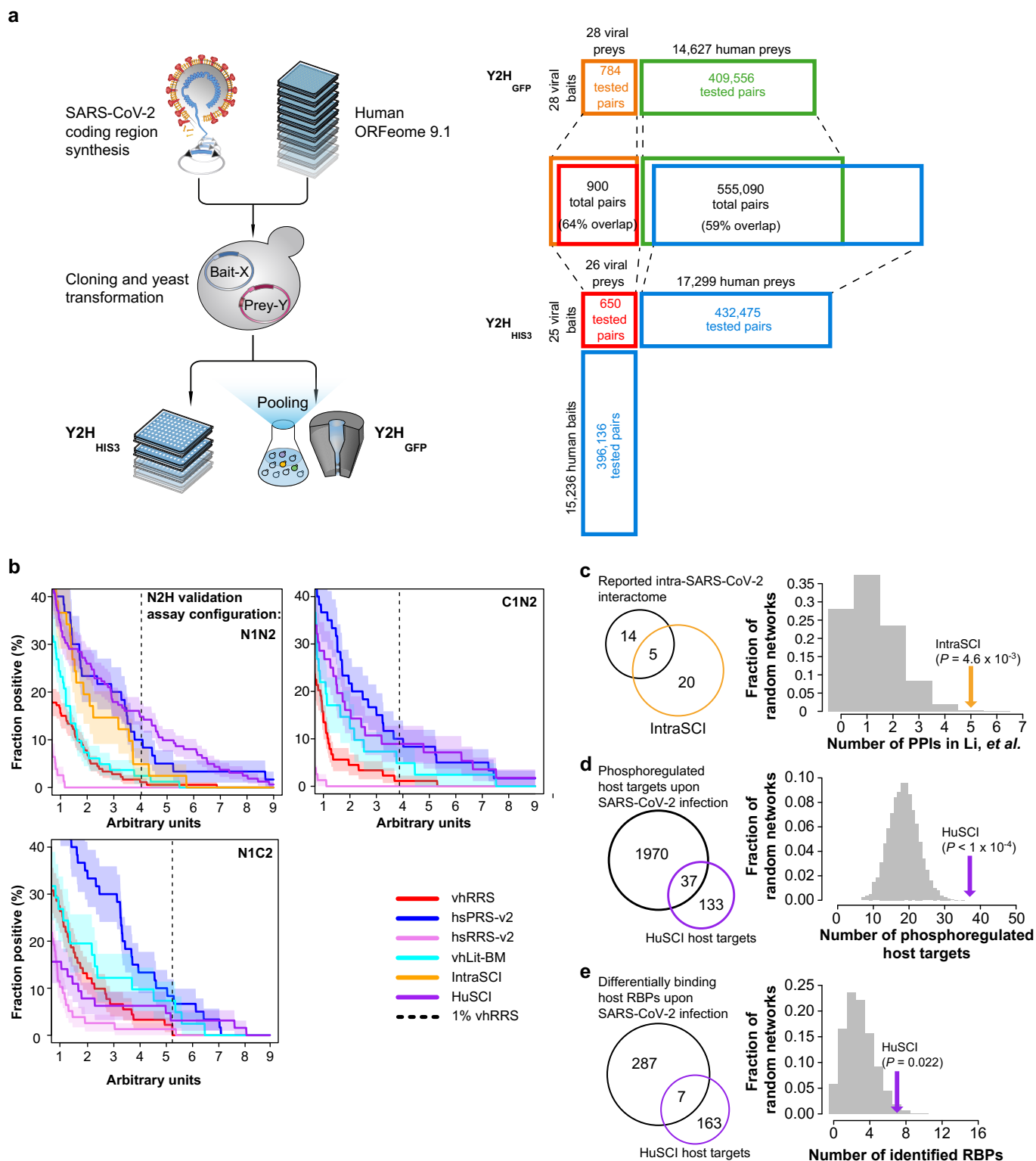
Extended Data Fig. 1 | Screening space, orthogonal validation and comparison of HuSCI and IntraSCI to previous SARS-CoV-2 related datasets. **a**, Schematic of the experimental contactome mapping pipelines (left) and screening space for each of the two parallel Y2H$_{HIS}$ and Y2H$_{GFP}$ screens (right). Proportional overlap is given relative to the union of protein pairs tested by both methods. **b**, Rate at which interactions are detected by yN2H for HuSCI and IntraSCI, as well as positive (hsPRS-v2 and vhLit-BM) and negative (hsRRS-v2 and vhRRS) benchmark sets, across stringency thresholds, error band: standard error of proportion. **c**, Left: overlap of previously identified intraviral SARS-CoV-2 interactions and IntraSCI; right: actual overlap (arrow) compared to n = 10,000 randomized control networks. One-sided, empirical P = 0.0046. **d**, Left: overlap of host targets identified in HuSCI and differentially phosphorylated proteins following infection by SARS-CoV-2; right: actual overlap (arrow) compared to n = 10,000 randomized control networks. One-sided, empirical P < 0.0001. **e**, Left: overlap of host targets identified in HuSCI and RNA Binding Proteins (RBPs) demonstrating differential RNA binding upon SARS-CoV-2 infection; right: actual overlap (arrow) compared to n = 10,000 randomized control networks. One-sided, empirical P = 0.022.

a

- ○ binary interactome
- ◌ AP-MS interactome
- ○ BioID interactome

b Fisher's exact test results for tissue bias comparisons relative to Human Protein Atlas

Data	P-value (fisher exact test)	P-value (adjusted, bonferroni)
HuSCI	0.0339	0.271
Gordon, et al.	3.92x10[-42]	3.13x10[-41]
Stukalov, et al.	1.81x10[-43]	1.45x10[-42]
Li, et al.	4.78x10[-23]	3.82x10[-22]
Nabeel, et al.	1.23x10[-26]	9.83x10[-26]
Laurent, et al.	1.19x10[-143]	9.52x10[-143]
St-Germain, et al.	2.41x10[-78]	1.93x10[-77]
Samavarchi-Tehrani, et al.	1.01x10[-136]	8.07x10[-136]

d

c Organotropism analysis of SARS-CoV-2 viral targets

Dataset	Airway	Brain	GI tract	Heart	Kidney	Liver	Lung	Muscle	Pancreas
HuSCI	4.12%	8.82%	4.12%	0.59%	1.18%	1.76%	0.00%	8.82%	1.76%
HPA	3.53%	13.15%	4.69%	1.97%	2.10%	4.76%	1.22%	5.18%	2.15%
Gordon, et al.	0.26%	2.34%	0.52%	0.78%	0.78%	4.69%	0.00%	3.12%	1.04%
Stukalov, et al.	1.71%	6.29%	3.66%	0.69%	2.51%	6.29%	0.46%	4.00%	1.14%
Li, et al.	1.40%	3.86%	0.35%	0.70%	0.70%	4.91%	0.35%	7.02%	1.05%
Nabeel, et al.	0.00%	3.25%	0.72%	1.44%	0.00%	2.17%	0.00%	6.14%	2.17%
Laurent, et al.	0.66%	7.00%	1.32%	0.94%	0.75%	3.15%	0.14%	4.51%	1.08%
St-Germain, et al.	0.79%	4.85%	1.98%	0.99%	0.69%	3.66%	0.20%	3.66%	1.29%
Samavarchi-Tehrani, et al.	1.12%	5.13%	1.96%	1.25%	0.94%	5.17%	0.36%	3.88%	1.29%

Extended Data Fig. 2 | Comparison of HuSCI with SARS-CoV-2 association and proximity datasets. a, Overlap of viral-human protein pairs between HuSCI, four AP-MS and three BioID based datasets (Gordon et al.[12,13], Stukalov et al.[9], Li et al.[14], Nabeel-Shah et al.[15], Laurent et al.[16], St-Germain et al.[17], Samavarchi-Tehrani et al.[18]). **b**, Statistical analysis of representation of host targets in common and specific expression groups from datasets in (**a**), compared to the Human Protein Atlas22 (HPA) (Fisher's exact test with Bonferroni correction). **c**, Organotropism analysis across SARS-CoV-2 infected organs from datasets in (**b**). The percentage of genes within each dataset with specific organotropism ('tissue-specific' expression in tissues grouped into organ systems). **b** and **c**, Full analysis is shown in Supplementary Table 4. **d**, Proportion of HuSCI host interactors per SARS-CoV-2 protein in which the human protein has: domains present in other interactors of the viral protein (shared); domains not present in other interactors of the viral protein (unique); no structural domains. Full analysis is shown in Supplementary Table 6.

a

Critical illness-associated loci and protein-encoding genes in HuRI

Study	SNP	Locus (bold: protein-encoding genes in HuRI)
Pairo-Castineira, E. et al., 2021	rs73064425	**LZTFL1**
	rs9380142	HLA-G
	rs143334143	**CCHCR1**
	rs3131294	NOTCH4
	rs10735079	**OAS1, OAS2**, OAS3
	rs2109069	DPP9
	rs74956615	**TYK2**
	rs2236757	IFNAR2
COVID-19 Host Genetics Initiative, 2021	rs10490770	**LZTFL1**
	rs10774671	**OAS1, OAS2**, OAS3
	rs77534576	**KAT7**, TAC4
	rs2109069	DPP9
	rs74956615	**ICAM1, ICAM3, ICAM4**, ICAM5, ZGLP1, **FDX2, RAVER1, TYK2**
	rs13050728	IFNAR2

b

c

d

e

f

Severe COVID-19 comorbidities and associated studies

Trait name	Phenotype	Type	Link to COVID-19 - summary	PMID
T2D_UKBS	TYPE-2-DIABETES	Cardiometabolic	Increased odds of in-hospital death with COVID-19, severity	32798472, 33661905, 33927066
BMIA	BMI	Anthropometric	High BMI associated with increased COVID-19 risk	33262790, 33661905
FAT_UKBS	BODY FAT PERCENTAGE	Anthropometric	Increased risk, related to BMI	33631142
HRET	RETICULOCYTE COUNT	Blood	RET count is marker for severe COVID-19	33178762
RET	RETICULOCYTE COUNT	Blood	RET count is marker for severe COVID-19	33178762
HC_UKBS	HIGH CHOLESTEROL	Cardiometabolic	High cholesterol associated with increased COVID-19 risk	33262790, 33308159
ADPN	ADIPONECTIN	Cardiometabolic	Linked to respiratory failure in COVID-19 patients	33904656
HYPOTHY_UKBS	HYPOTHYROIDISM	Endocrine system disease	Associated with prolonged COVID-19 induced anosmia; can be caused by COVID-19	33879534, 34094769, 32436948
SCZ_UKBS	SCHIZOPHRENIA	Psychiatric-neurologic	Greater risk of adverse outcomes from COVID-19	34079487, 33502436
GIANT_HIP	HIP CIRCUMFERENCE	Anthropometric	Increased risk, not significant after accounting for BMI	33661905, 32835759
IBD_UKBS	INFLAMMATORY BOWEL DISEASE	Immune	Possible immunologic link	33079178
OST_UKBS	OSTEOPOROSIS	Skeletal system disease	Possible immunologic link	28930556
EGG_PHF	PUBERTAL GROWTH	Anthropometric		
GIANT_HEIGHT	BODY HEIGHT	Anthropometric		
NEUROT_UKB	NEUROTICISM	Psychiatric_neurologic		

g

Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Traits associated with COVID-19 severity. a, Table showing COVID-19 critical illness associated loci from two GWAS meta-analyses[32,33]. Locus-associated proteins present in HuRI are marked in bold. **b**, Genes in indicated COVID-19 datasets ranked across the human genome by number of publications. Error bars are 95% confidence intervals of the mean, calculated by 1,000 bootstrap samples (from top to bottom n = 45, 170, 383, 876, 29, 75, 25, 71, 233, 46, 49, 45, 15, 33, 10, 71, 97, 47, 58, 9, 46, 22, 20, 23, 39). **c**, Virus-interactor enrichment in contactome: number of direct SARS-CoV-2 protein interacting HuSCI proteins in a HuRI subnetwork formed by proteins encoded by COVID-19 critical illness associated loci[33] (marked in bold in table (**a**)) and their first level interactors (arrow) compared to n = 10,000 randomized control networks (gray distribution). One-sided, empirical *P* = 0.012. **d**, Virus-interactor enrichment in co-complex associations: number of SARS-CoV-2 associated human proteins in two AP-MS based studies[9,13] in a subnetwork formed by proteins encoded by COVID-19 critical illness associated loci[32,33] (marked in bold in table in (**a**)) and their first level interactors (arrow) either in HuRI or BioPlex 3.0. The comparisons are against n = 10,000 randomized control networks (gray distribution). One-sided, empirical *P* values are shown for each dataset. **e**, Upset plots showing number of communities targeted by SARS-CoV-2 (left) and associated with severe COVID-19 (right) in HuSCI and AP-MS based datasets. **f**, Table showing 15 traits for genetic variation identified within targeted network communities. An association with severe COVID-19 comorbidities is indicated, as well as trait references: T2D_UKBS[91–93], BMIA[92,94], FAT_UKBS[95], HRET[96], RET[96], HC_UKBS[94,97], ADPN[98], HYPOTHY_UKBS[43,99,100], SCZ_UKBS[44,101], GIANT_HIP[92,102], IBD_UKBS[103], OST_UKBS[104], EGG_PHF, GIANT_HEIGHT, NEUROT_UKB. **g**, Grouping of 31 network communities with significantly associated traits shown in Fig. 3d by protein membership measured by Jaccard similarity according to legend.

a Summary statistics for unstimulated NF-κB reporter induction.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value
EV (pMH) vs. IKK	-0.6707	-0.8554 to -0.4859	Yes	****	<0.0001
EV (pMH) vs. A20	0.01772	-0.1670 to 0.2025	No	ns	0.9996
EV (pMH) vs. E	-0.01174	-0.1965 to 0.1730	No	ns	0.9997
EV (pMH) vs. N	0.01194	-0.1728 to 0.1967	No	ns	0.9997
EV (pMH) vs. NSP9	-0.01851	-0.2033 to 0.1663	No	ns	0.9996
EV (pMH) vs. NSP14	-0.1919	-0.3767 to -0.007137	Yes	*	0.0395
EV (pMH) vs. NSP16	-0.005107	-0.1899 to 0.1797	No	ns	0.9999
EV (pMH) vs. NSP6	-0.06448	-0.2492 to 0.1203	No	ns	0.8862
EV (pMH) vs. NSP7	-0.1404	-0.3252 to 0.04431	No	ns	0.1946

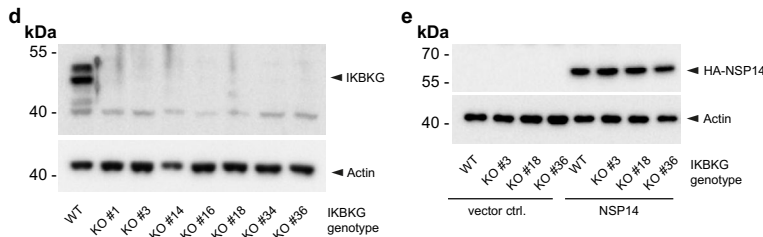Summary statistics for TNFα stimulated NF-κB reporter induction.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value
EV (pMH) vs. IKK	-0.4133	-0.9394 to 0.1128	No	ns	0.1705
EV (pMH) vs. A20	0.4687	-0.05741 to 0.9948	No	ns	0.0953
EV (pMH) vs. E	-0.1383	-0.6644 to 0.3878	No	ns	0.9729
EV (pMH) vs. N	-0.1694	-0.6955 to 0.3567	No	ns	0.9213
EV (pMH) vs. NSP9	-0.2245	-0.7506 to 0.3016	No	ns	0.7529
EV (pMH) vs. NSP14	-0.7203	-1.246 to -0.1942	Yes	**	0.0047
EV (pMH) vs. NSP16	-0.0223	-0.5484 to 0.5038	No	ns	0.9998
EV (pMH) vs. NSP6	-0.4657	-0.9917 to 0.06044	No	ns	0.0985
EV (pMH) vs. NSP7	-0.4627	-0.9888 to 0.06335	No	ns	0.1016

b Summary statistics for unstimulated and TNFα stimulated NF-κB reporter induction.

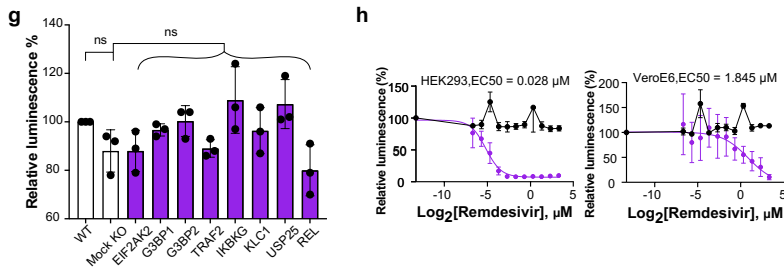|	Unstimulated					TNFα stimulated			
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value
EV vs. NSP14 - 2 ug	0.09108	-0.8869 to 1.069	No	ns	0.9996	0.5681	-0.4894 to 1.625	No	ns	0.5762
EV vs. NSP14 - 4 u	-1.1390	-2.117 to -0.1606	Yes	*	0.0183	0.3376	-0.7198 to 1.395	No	ns	0.946
EV vs. NSP14 - 6 u	-3.7360	-4.714 to -2.758	Yes	****	<0.0001	-1.564	-2.621 to -0.5061	Yes	**	0.0012
EV vs. NSP16 - 2 ug	0.2756	-0.7024 to 1.254	No	ns	0.9570	0.4327	-0.6247 to 1.490	No	ns	0.8267
EV vs. NSP16 - 4 ug	0.4618	-0.5161 to 1.440	No	ns	0.6528	0.2636	-0.7938 to 1.321	No	ns	0.9882
EV vs. NSP16 - 6 ug	0.4343	-0.5437 to 1.412	No	ns	0.7098	0.06232	-0.9951 to 1.120	No	ns	0.9998
EV vs. NSP9 - 2 ug	0.3445	-0.6334 to 1.322	No	ns	0.8752	0.3243	-0.7331 to 1.382	No	ns	0.9566
EV vs. NSP9 - 4 ug	0.4044	-0.5735 to 1.382	No	ns	0.7696	0.08329	-0.9741 to 1.141	No	ns	0.9997
EV vs. NSP9 - 6 ug	0.4527	-0.5252 to 1.431	No	ns	0.6716	0.1077	-0.9497 to 1.165	No	ns	0.9996

c Summary statistics for NF-κB transcriptional reporter activity under unstimulated (left), TNFα stimulated (middle) and NSP14 induced conditions in WT and IKBKG-KO HEK293 cells.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value
vector ctrl.					
WT vs. IKBKG KO #3	0.2456	-0.2774 to 0.7685	No	ns	0.5134
WT vs. IKBKG KO #18	0.2625	-0.2604 to 0.7855	No	ns	0.4622
WT vs. IKBKG KO #36	0.195	-0.3280 to 0.7179	No	ns	0.6755
vector ctrl. + TNFα					
WT vs. IKBKG KO #3	3.984	3.461 to 4.507	Yes	****	<0.0001
WT vs. IKBKG KO #18	3.994	3.471 to 4.517	Yes	****	<0.0001
WT vs. IKBKG KO #36	3.907	3.384 to 4.429	Yes	****	<0.0001
6 ug NSP14					
WT vs. IKBKG KO #3	4.654	4.131 to 5.177	Yes	****	<0.0001
WT vs. IKBKG KO #18	4.882	4.359 to 5.405	Yes	****	<0.0001
WT vs. IKBKG KO #36	4.321	3.798 to 4.844	Yes	****	<0.0001

d Western blot: kDa markers 55, 40, 40; IKBKG, Actin. Lanes WT, KO #1, KO #3, KO #14, KO #16, KO #18, KO #34, KO #36. IKBKG genotype.

e Western blot: kDa markers 70, 55, 40; HA-NSP14, Actin. Lanes WT, KO #3, KO #18, KO #36 (vector ctrl.) and WT, KO #3, KO #18, KO #36 (NSP14). IKBKG genotype.

f Summary statistics for viral replication assay.

Dunn's multiple comparisons test	Mean rank diff.	Significant?	Summary	Adjusted P Value
Mock KO vs. WT	3.667	No	ns	>0.9999
Mock KO vs. G3BP2	29.28	No	ns	0.1564
Mock KO vs. G3BP1	24.11	No	ns	0.4512
Mock KO vs. USP25	68.11	Yes	****	<0.0001
Mock KO vs. TRAF2	36.00	Yes	*	0.031
Mock KO vs. REL	71.72	Yes	****	<0.0001
Mock KO vs. IKBKG	42.67	Yes	**	0.0047
Mock KO vs. KLC1	50.67	Yes	***	0.0003
Mock KO vs. EIF2AK2	15.44	No	ns	>0.9999

g Bar graph: Relative luminescence % (y-axis, 60–140). Categories: WT, Mock KO, EIF2AK2, G3BP1, G3BP2, TRAF2, IKBKG, KLC1, USP25, REL. ns annotations.

h Dose-response curves: HEK293, EC50 = 0.028 µM; VeroE6, EC50 = 1.845 µM. X-axis Log$_2$[Remdesivir], µM; Y-axis Relative luminescence (%).

Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Effect of viral proteins on NF-κB reporter activity and of viral interactors on viral replication. a, Tables showing statistical details of NF-κB transcriptional reporter activity in the absence and presence of selected viral proteins under unstimulated (top) and TNFα stimulated (bottom) conditions. One-way ANOVA with Dunnett's multiple comparisons test, n = 3, adjusted *P* values are shown. **b**, Table showing statistical details of NF-κB transcriptional reporter activity at different amounts of transfected viral protein-encoded plasmid under unstimulated (left) and TNFα stimulated conditions (right). One-way ANOVA with Dunnett's multiple comparisons test, n = 3 and n = 6, respectively, adjusted *P* values are shown. **a** and **b**, Raw data and full analysis is shown in Supplementary Table 9. **c**, Table showing statistical details of NF-κB transcriptional reporter activity under unstimulated (left), TNFα-stimulated (middle) and NSP14-induced conditions in WT and IKBKG KO HEK293 cells (two-way ANOVA with Dunnett's multiple comparisons test, n = 3), adjusted *P* values are shown. **d**, Representative anti-IKBKG (top) western blot demonstrating levels of IKBKG in WT and three independent IKBKG knockout clones of HEK293 cells

relative to actin beta (ACTB) loading controls (bottom). **e**, Representative anti-hemagglutinin (HA) western blot demonstrating levels of tagged NSP14 protein in NF-κB induction experiments relative to actin beta (ACTB) loading controls (bottom). **f**, Table showing statistical details of viral replication in wild-type, mock KO and CRISPR KOs of the indicated HuSCI host proteins. Kruskal-Wallis with Dunn's multiple comparisons test, n = 9. Adjusted *P* values are shown. **g**, Cell viability of mock KO and CRISPR KOs of the indicated HuSCI host proteins relative to WT cells. Kruskal-Wallis with Dunn's multiple comparisons test, n = 3. Adjusted, Fisher's exact *P* values are shown. **f** and **g**, Raw data, Fisher's exact *P* values, and full analysis is shown in Supplementary Table 10. **h**, Cell viability and relative replication of icSARS-CoV-2-nanoluciferase in HEK293 cells (left) and Vero E6 cells (right) at different concentrations of remdesivir. The EC50 values shown for each cell line were calculated with a variable slope model. Error bars: standard deviation of the mean, n = 3 biological repeats, full analysis in Supplementary Table 11.

a

AD-N β | AD-N γ | AD-N α | AD-N W.Hu1

b

AD-E β | AD-E W.Hu1

c

AD-empty
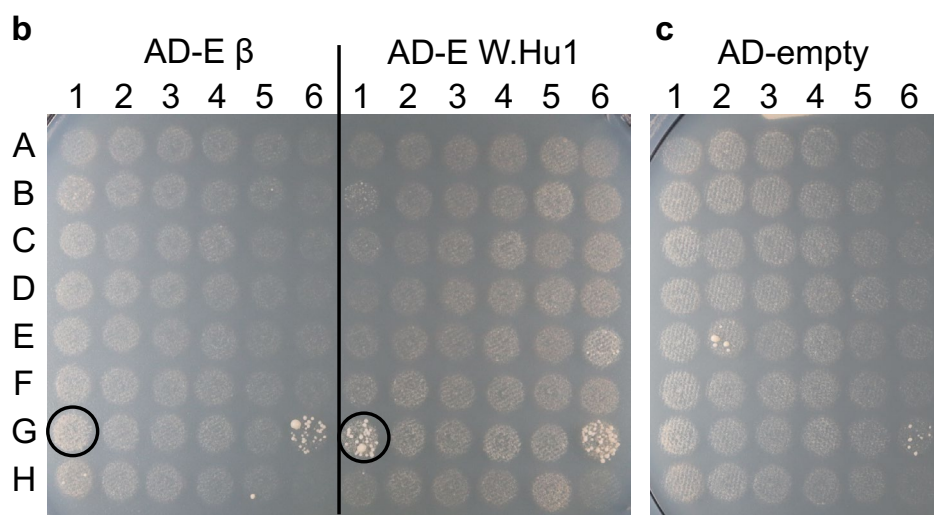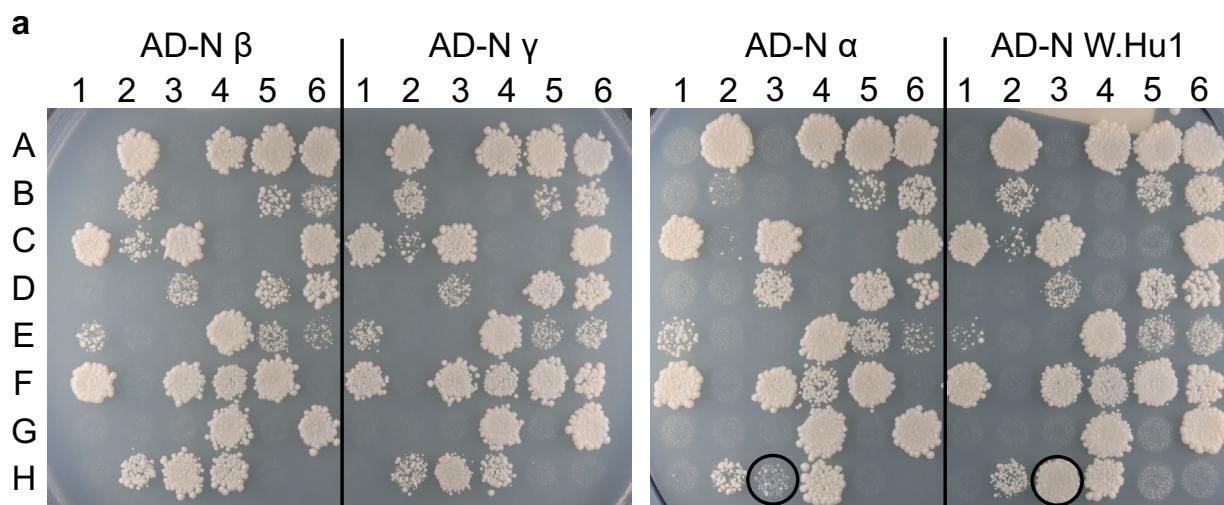
d

DB-x

	1	2	3	4	5	6
A	N/A	NKRF	N/A	G3BP2	CDC37	ELOA
B	N/A	N/A	N/A	N/A	N/A	EIF2AK2
C	MORF4L1	N/A	KLC1	N/A	N/A	ADARB1
D	N/A	N/A	N/A	N/A	TARBP2	KLC2
E	BYSL	N/A	N/A	NTAQ1	LYAR	N/A
F	EHHADH	N/A	PRKRA	KLC4	G3BP1	RAD23B*
G	BAG4	N/A	N/A	USO1	N/A	USP9X
H	N/A	MORF4L2	ARPC3	N/A	N/A	DB-empty

Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Mutations of SARS-CoV-2 variants affect specific interactions with uSCI host targets. a, Y2H$_{HIS3}$ yeast growth on selective plates of HuSCI interaction partners as DB-fusion proteins tested against AD-fusion of the SARS-CoV-2 Nucleocapsid (AD-N) protein (Wuhan-Hu1, original screen) and AD-N containing 'lineage defining' amino acid substitutions: D3L and S235F (α-strain), T205I (β-strain), or P80R (γ-strain). Shown is one representative result of 5 repeats. **b**, Y2H$_{HIS3}$ yeast growth on selective plates of HuSCI interaction partners as DB-fusion proteins tested against AD-fusion of the SARS-CoV-2

Envelope (AD-E) protein (Wuhan-Hu1, original screen) or AD-E containing 'lineage defining' substitution P71L (β-strain). Shown is one representative Y2H$_{HIS3}$ result on selective media, out of 2 repeats. **a** - **c**, Black circles indicate changes in yeast colony growth between human proteins tested against viral variant ORFs or the originally screened Wuhan strain ORFs observed consistently across all repeats. **c**, AD-empty control plate for **a, b** indicates lack of autoactivation. **d**, Layout of DB-fusion HuSCI interactors (purple) tested with AD fusion SARS-CoV-2 proteins or AD-empty control, respectively in **a** - **c**. N/A indicates human interactors.

nature portfolio

Corresponding author(s): Caroline Demeret, Marc Vidal, Michael A. Calderwood, Frederick P. Roth and Pascal Falter-Braun

Last updated by author(s): Aug 8, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Bowtie2 2.2.3
	Illumina bcl2fastq 2.20.0.422
	Python 2.7.10
	Custom code for Y2H screening pipeline
	Tissue Atlas dataset from the Human Protein Atlas database (version: 2021.04.09)

Data analysis	Cytoscape 3.8.2
	Matlab 2016a
	gProfiler (Database versions: Ensembl 104, Ensembl Genomes 51 and Wormbase ParaSite 15)
	R 3.6.1
	linkcomm R package
	Metascape web-platform
	Pfam release 34.0
	FlowJo Software 10
	MAGMA v1.09a
	MikroWin 2010
	Graphpad Prism v. 7.04,
	Graphpad Prism v. 9,
	AmiGO 2

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our <u>policy</u>

The PPI data from this publication have been submitted to the IMEx (http://www.imexconsortium.org) consortium through IntAct and assigned the identifier IM-28880104. All data from the study are included in the manuscript and associated files. Source data are provided with this paper.
The following data were obtained from the respective original publications: phosphorylation changes upon SARS-CoV-2 infection[10,11]; RNA-binding changes upon SARS-CoV-2 infection[12]; AP-MS virus-host association data: Gordon et al.[14,15], Stukalov et al.[10], Li et al.[13], Nabeel-Shah et al.[16]; BioID virus-host proximity data: Laurent et al.[17], St-Germain et al.[19] Samavarchi-Tehrani et al.[18]; Human expression data: Human Proteome Atlas[22], SARS-CoV-2 organotropism[23,99]; human host interactome: HuRI[27]; GWAS data for severe COVID-19 illness[33,34];
GWAS summary statistics for 114 traits: doi:10.5281/ZENODO.3518299; Interaction data for other viruses were downloaded from IntAct[9] (version: April 28, 2020). Publication counts downloaded from gene2pubmed (NCBI), on 2021-11-16.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see <u>nature.com/documents/nr-reporting-summary-flat.pdf</u>

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for PPI validation assay experiments was chosen based on a 95% CI value of 5% inputting and an expected recovery rate of 20%, from previous similar experiments, using the standard error of proportion. For IntraSCI the full dataset was tested. For all cell biological assays 3 - 8 biological repeats were done. Many years of experience have convinced us that this produces robust results.
Data exclusions	No data were excluded.
Replication	All attempts at replication were successful. For Y2H, 3 fold-replication is part of the standard quality control pipeline. In all cell biology experimnets, i.e. viral replication assays and NF-kB reporter activation assays, biological results were replicated at least twice (3 biological repeats).
Randomization	Randomization for network permutations was done while maintaining network topology (maintaining the number of interactions but shuffling the edges). yN2H validations were done with a fully randomized setup to avoid experimental bias and eliminate covariates, i.e. all pairs, except hsPRSv2/hs/RRv2, were randomly distributed across the plates. For experimental reasons, and because extensive prior data are available for hsPRSv2/hs/RRv2 benchmark pairs these were randomly mixed among each other, but separate from other benchmark and test pairs. For all cell culture and validation experiments, all samples were treated by the same researcher in parallel in random order to eliminate covariates. In all experiments, negative controls from the same experimental batch were used as initial reference.
Blinding	Y2H screens were done in a fully random and unbiased manner. Manual scoring of Y2H pairwise tests was blinded. Experimental design of yN2H validation was blinded during the experiment. For cell culture experiments with small sample sizes blinding was not possible, but all experiments are internally controlled to reduce researcher biased experimental differences.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	1. Anti-b-Actin (SCBT, sc-47778, RRID: AB_2714189) 2. anti-FLAG M2 (Sigma Aldrich, F3165, RRID:AB_259529) 3. anti-HA (Sigma-Aldrich, 11583816001, RRID:AB_514505) 4. Secondary antibody (JacksonImmuno Research, Jim-715-035-150)
Validation	The antibodies have been validated by the manufacturers and were not revalidated in this study.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	1. A549 (purchased from ATCC) 2. Hek293 (RRID: CVCL_0045, purchased from DSMZ) 3. Vero E6 (ATCC CRL-1586), kindly provided by Prof. Nicolas Gillet (University of Namur, Belgium) 4. Yeast strains Y8800 and Y8930: the same strains used in the many previous studies (Rolland, et al. Cell. 2014 Nov 20;159(5):1212-1226.) 5. Yeast strains - RY3011 and RY3031: Derived from Y8800 and Y8930 in Dr. Fritz Roth's lab at the University of Toronto
Authentication	Vero E6 cells were re-authenticated by RNA-Seq. Other cell lines were purchased from either ATCC or DSMZ and not re-authenticated.
Mycoplasma contamination	All cell lines were tested by PCR for mycoplasma contamination and found to be negative
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.