

1 **Human phospho-signaling networks of SARS-CoV-2 infection are rewired by population**
2 **genetic variants**

3

4 Diogo Pellegrina^{1,*}, Alexander T. Bahcheli^{1,2,*}, Michal Krassowski³, Jüri Reimand^{1,2,4,@}

5

6 1. Computational Biology Program, Ontario Institute for Cancer Research, Toronto, ON, Canada

7 2. Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

8 3. Nuffield Department of Women's and Reproductive Health, Medical Sciences Division,
9 University of Oxford, United Kingdom

10 4. Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

11 * These authors contributed equally to the study

12 @ correspondence: Juri.Reimand@utoronto.ca

13 **ABSTRACT**

14 **SARS-CoV-2 infection hijacks signaling pathways and induces protein-protein interactions**
15 **between human and viral proteins. Human genetic variation may impact SARS-CoV-2**
16 **infection and COVID-19 pathology; however, the role of genetic variation in these signaling**
17 **networks remains uncharacterized. We studied human single nucleotide variants (SNVs)**
18 **affecting phosphorylation sites modulated by SARS-CoV-2 infection, using machine**
19 **learning to identify amino acid changes altering kinase-bound sequence motifs. We found**
20 **2033 infrequent phosphorylation-associated SNVs (pSNVs) that are enriched in sequence**
21 **motif alterations, potentially reflecting the evolution of signaling networks regulating host**
22 **defenses. Proteins with pSNVs are involved in viral life cycle processes and host responses,**
23 **including regulators of RNA splicing and interferon response, as well as glucose**
24 **homeostasis pathways with potential associations with COVID-19 co-morbidities. Certain**
25 **pSNVs disrupt CDK and MAPK substrate motifs and replace these with motifs recognized**
26 **by Tank Binding Kinase 1 (TBK1) involved in innate immune responses, indicating**
27 **consistent rewiring of infection signaling networks. Our analysis highlights potential**
28 **genetic factors contributing to the variation of SARS-CoV-2 infection and COVID-19 and**
29 **suggests leads for mechanistic and translational studies.**

30 INTRODUCTION

31 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the coronavirus disease
32 2019 (COVID-19) pandemic has caused millions of deaths worldwide and continues to evolve as
33 more pathogenic and infectious variants of the virus emerge. The clinical manifestations and
34 outcomes of COVID-19 are complex, ranging from asymptomatic infection to fatal respiratory
35 and multi-organ failure, as well as long-term symptoms after recovery. Risk factors of severe
36 disease include advanced age, a weakened immune system, and pre-existing health conditions
37 such as hypertension, diabetes, and obesity ¹. Ethnic and demographic patient characteristics that
38 are at least partially explained by socio-economic factors also affect disease risk ²⁻⁴. Recent
39 genome-wide association studies (GWAS) of COVID-19 have shed light on human genetic
40 variation contributing to SARS-CoV-2 disease burden and mortality rates, highlighting genes
41 associated with ABO blood groups, antiviral pathways and tyrosine kinase signaling ⁵⁻⁷.
42 However, GWAS findings often occur in intergenic regions where molecular mechanisms
43 remain elusive, and rare variants potentially contributing to disease are challenging to detect.
44 Thus, additional work is needed to find and interpret genetic variants contributing to SARS-
45 CoV-2 infection and COVID-19 pathology.

46 Molecular interaction networks are perturbed by host-pathogen interactions and host disease
47 mutations that either disable proteins in the networks or alter their interactions ⁸. Phosphorylation
48 is a key component of cellular signaling networks that acts as a reversible molecular switch
49 controlling protein function and interactions. This post-translational modification (PTM) is
50 conducted by kinases that recognize sequence motifs at protein phosphorylation sites (*i.e.*,
51 phosphosites). SARS-CoV-2 infection alters phosphorylation networks in host cells by
52 promoting casein kinase 2 (CK2) and mitogen-activated protein kinase (MAPK) pathways and
53 inhibiting mitotic kinases, resulting in cell cycle arrest and cytoskeletal changes to favor virus
54 pathology ⁹. Phosphorylation networks also control host anti-viral and immune responses to
55 SARS-CoV-2 infection. For example, the TANK Binding Kinase 1 (TBK1) and IKK-epsilon
56 kinase of the NF- κ B pathway initiate innate antiviral response by phosphorylating interferon
57 regulatory factors (IRF) that regulate interferon genes ^{10,11}. Interferons activate JAK-STAT, p38
58 MAPK, and PI3K/Akt signaling pathways that are dysregulated in SARS-CoV-2 infection and
59 COVID-19 ^{9,12,13}. SARS-CoV-2 proteins bind TBK1 to suppress host immune responses ^{14,15}.

60 NF- κ B hyperactivation has been associated with cytokine storms in COVID-19 where excessive
61 production of proinflammatory cytokines can cause fatal damage to the host ¹⁶. Besides host
62 protein, post-translational modifications of virus proteins may also contribute to the complexity
63 of infection and host anti-viral responses ¹⁷.

64 Genetic variants involved in human disease are known to affect kinase signaling networks. For
65 instance, inherited disease mutations and somatic mutations in cancer driver genes frequently
66 erase phosphosites or alter kinase binding motifs, potentially causing rewiring of kinase
67 signaling ¹⁸⁻²¹. Conversely, phosphosites tend to have lower genetic variation in the human
68 population, underlining the importance of conserved phosphosites in evolution ^{18,22}. This
69 suggests that human genetic variation of signaling networks responding to the novel SARS-CoV-
70 2 infection may contribute to a range of symptoms, disease severity, and long-term outcomes of
71 COVID-19 patients.

72 We hypothesized that genetic variation in phosphosites differentially phosphorylated in SARS-
73 CoV-2 infection can alter kinase signaling interactions and thereby reveal mechanistic insights
74 into the genes and pathways involved in infection and disease. We studied the gnomAD dataset
75 ²³, the largest uniformly processed map of human genetic variation comprising exome
76 sequencing data of more than 100,000 individuals, and mapped missense single nucleotide
77 variants (SNVs) to human protein phosphosites responding to SARS-CoV-2 infection ⁹. With a
78 machine learning approach, we uncovered hundreds of phosphorylation-associated SNVs
79 (pSNVs) that modify kinase-bound sequence motifs and potentially rewire kinase-substrate
80 interactions. Our study helps decipher the role of human genome variation in virus responses and
81 disease outcomes and enables advances in therapy and biomarker development.

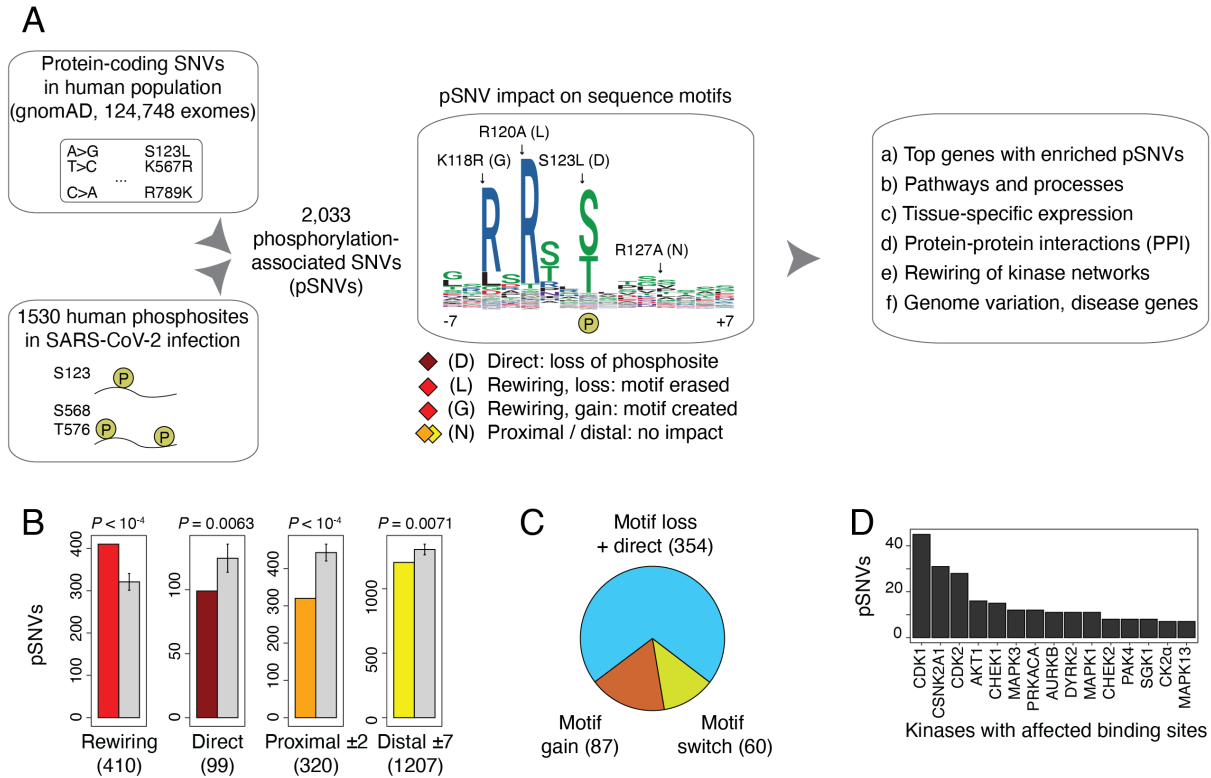


Figure 1. Phosphorylation-associated single nucleotide variants (pSNVs) in human phosphosites of SARS-CoV-2 infection. **A.** Overview of analysis. Missense SNVs in the human population (allele frequency $AF_{popmax} \geq 10^{-4}$ in gnomAD) were mapped to host protein phosphosites significantly differentially phosphorylated in SARS-CoV-2 infection⁹. Sequence motif analysis was used to evaluate pSNV impact to predict gains and losses of kinase binding sites. Genes, pathways, and molecular interaction networks with frequent pSNVs were analyzed. **B.** pSNVs of four functional classes were studied: (1) motif-rewiring pSNVs altering sequence motifs in kinase binding sites (MIMP posterior $prob > 0.5$), (2) direct pSNVs altering phosphorylated residues, and (3-4) pSNVs with no motif-rewiring predictions proximal and distal to phosphosites ($\pm 1-2$ and $\pm 3-7$ residues, respectively). Expected numbers of pSNVs (grey) represent control phosphosites sampled randomly from the proteome. Empirical P-values and pSNV counts are shown. pSNVs predicted to cause rewiring of sequence motifs are enriched compared to controls (shown in red). **C.** Impact of motif-rewiring pSNVs on protein phosphorylation. Pie chart shows pSNVs predicted to cause loss of phosphosites (*i.e.*, pSNVs substituting phospho-residues and/or disrupting sequence motifs; blue), gain of phosphosites (*i.e.*, pSNVs inducing new sequence motifs; orange), and motif switches (pSNVs erasing one motif and inducing another motif). **D.** A subset of pSNVs occur at high-confidence binding sites of kinases. Bar plot shows the top 15 kinases whose binding sites have the largest numbers of pSNVs.

82

83 **RESULTS**

84 SARS-CoV-2 associated phosphosites are enriched in pSNVs rewiring kinase binding 85 motifs

86 To examine the human genetic variation in signaling networks responding to SARS-CoV-2
87 infection, we studied exome sequencing data of 124,748 individuals from 16 populations in the
88 gnomAD dataset ²³, focusing on 1,111,194 amino acid substitutions that were observed at least
89 once per 10,000 individuals in at least one of the populations ($AF_{popmax} > 10^{-4}$). We defined
90 phosphorylation-associated SNVs (pSNVs) as missense SNVs (*i.e.*, amino acid substitutions)
91 that affected 1530 host phosphosites differentially phosphorylated in SARS-CoV-2 infection
92 based on an earlier phosphoproteomics study ($FDR < 0.05$; 24h timepoint) ⁹. We mapped pSNVs
93 in flanking windows of ± 7 amino acid residues around phosphorylated residues to study
94 sequence motifs bound by kinases (**Figure 1A**). To predict the functional impact of pSNVs on
95 protein phosphorylation, we evaluated the effects of pSNVs on 125 types of sequence motifs
96 bound by kinases of 77 families. We used MIMP ²⁴, a machine learning method trained on
97 kinase binding sequences, that evaluates whether a pSNV disrupts an existing motif or creates a
98 new motif relative to the reference protein sequence. Higher Bayesian posterior probabilities
99 computed by MIMP reflect an increased likelihood of pSNVs causing rewiring of sequence
100 motifs.

101 We found 2,033 pSNVs in 987 SARS-CoV-2 -associated phosphosites and 693 genes and
102 assigned these to four classes based on their predicted functional impact (**Figure 1B**,
103 **Supplementary Table 1**). First, *direct* pSNVs (99 or 5%) replaced central phospho-residues and
104 caused loss of phosphosites, likely representing the pSNV class of highest impact. Second, *motif-*
105 *rewiring* pSNVs (410 or 20%) created or disrupted sequence motifs, potentially causing gains or
106 losses kinase binding at the phosphosites (MIMP posterior *prob* > 0.5). The two remaining
107 classes of pSNVs lacked functional predictions based on sequence analysis and included 1527
108 (75%) pSNVs annotated as *proximal* or *distal* to the closest phosphosite. When combining direct
109 and motif-rewiring pSNVs, loss of phosphorylation through motif disruption or phospho-residue
110 replacement was predicted for 354 pSNVs. 87 pSNVs induced new sequence motifs at SARS-
111 CoV-2-associated phosphosites and were predicted to cause gain of phosphorylation (**Figure 1C**,
112 **Supplementary Table 2**). Interestingly, 60 pSNVs caused motif switching where the pSNV
113 simultaneously disrupted one motif and replaced it with another motif at the same phosphosite.

114 Overall, hundreds of pSNVs in the human population are predicted to alter SARS-CoV-2 driven
115 signaling in host cells through the range of kinase sequence motifs we studied.

116 We evaluated the statistical significance of these potentially functional pSNVs by randomly
117 sampling human phosphosites as controls. Compared to experimentally determined phosphosites
118 in the human proteome, SARS-CoV-2 associated phosphosites were enriched in motif-rewiring
119 pSNVs (410 pSNVs observed vs. 321 ± 21 pSNVs expected; $P < 10^{-4}$) (**Figure 1B**). In contrast,
120 direct pSNVs replacing central phospho-residues, as well as proximal and distal pSNVs lacking
121 motif-based predictions were found less frequently in SARS-CoV-2 associated phosphosites.
122 Some pSNVs (213 or 10%) occurred at well-defined phosphosites where the phosphorylating
123 kinases have been identified previously, such as cyclin dependent kinases, casein 2 kinases,
124 checkpoint kinases, and MAP kinases (**Figure 1D**), in agreement with studies of the human
125 signaling network modulated upon SARS-CoV-2 infection⁹ and adding confidence to our
126 sequence-based predictions of pSNV impact. Enrichment of motif-rewiring pSNVs suggests that
127 the signaling network responding to SARS-CoV-2 infection in humans may vary structurally
128 such that certain kinase-substrate interactions are gained or lost some individuals. This may
129 reflect positive evolutionary selection in adaptation to viral infection and potentially cause
130 variation in SARS-CoV-2 infection and COVID-19 disease course.

131

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

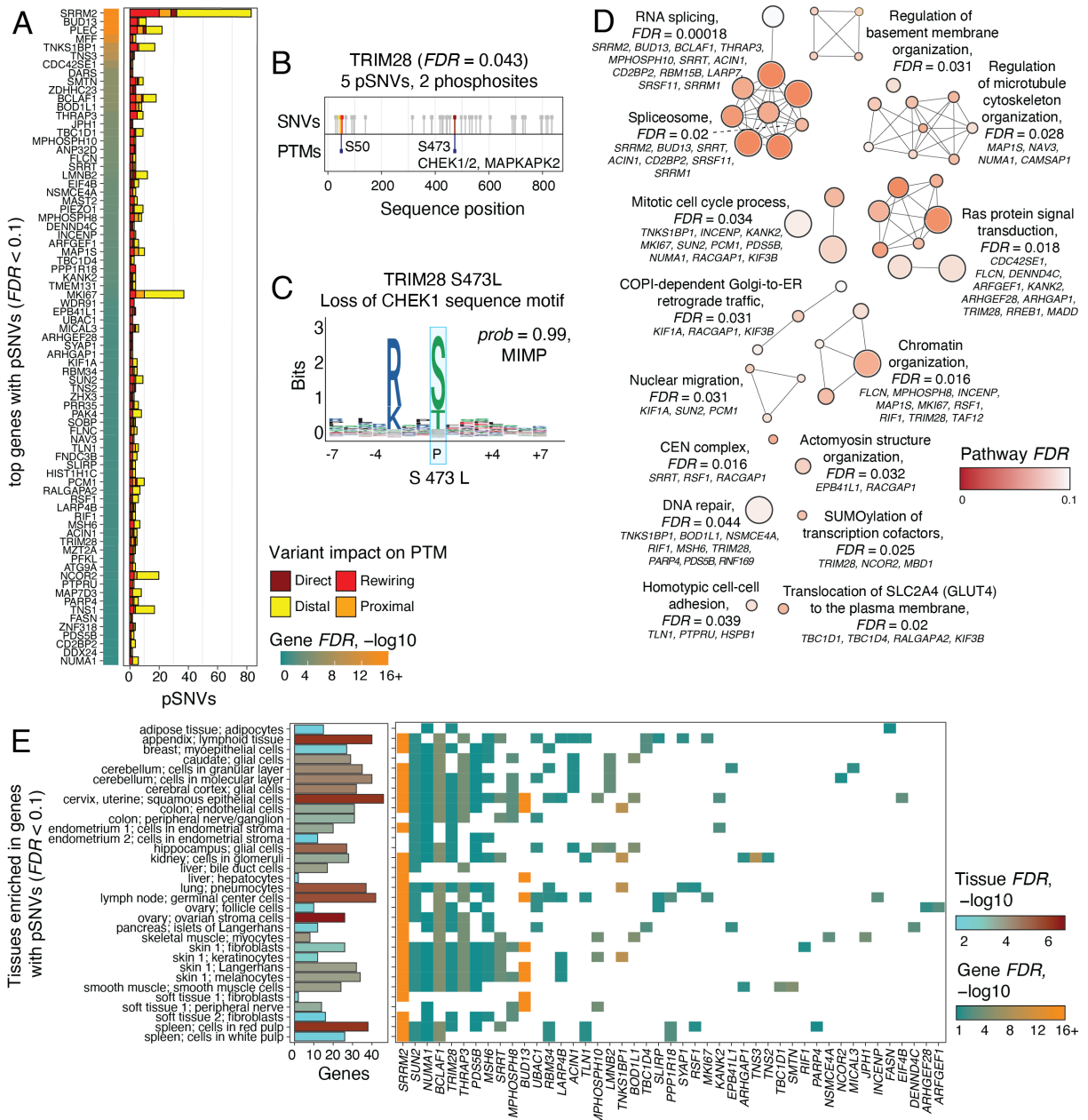


Figure 2. Top genes, pathways, and tissue-specific expression patterns of SARS-CoV-2 associated pSNVs.

A. Top genes with pSNVs predicted to alter phosphosites modulated by SARS-CoV-2 infection ($FDR < 0.1$).

Genes are scored by the probability of at least one pSNV altering SARS-CoV-2-associated phosphosites. **B.**

Example of a top gene with pSNVs: TRIM28, a transcriptional and epigenetic regulator involved in innate

immune response, includes five pSNVs in two phosphosites: S473 phosphorylated by CHEK and MAPKAP2

kinases, and the phosphosite S50 with no known kinase. **C.** The pSNV S473L in TRIM28 substitutes the

phospho-residue and disrupts a CHEK1 sequence motif with a high-confidence prediction of motif alteration

(MIMP posterior $prob > 0.99$). the pSNV potentially alters the regulation of interferons by TRIM28. **D.** pSNVs

are enriched in biological processes and molecular pathways (ActivePathways, $FDR < 0.1$). The enrichment map is a network of pathways and processes as nodes that are connected by edges if the pathways share many genes. Each subnetwork represents a distinct functional theme. Top genes with pSNVs are listed for each theme and include additional genes found in the pathway analysis (gene $FDR < 0.25$). **E.** Tissue-specific expression of genes with frequent pSNVs, as detected by enrichment analysis of Human Protein Atlas gene sets (ActivePathways, $FDR < 0.1$). Bar plot (left) shows all the genes with pSNVs expressed in each tissue. Grid plot (right) shows the tissues where the top genes with pSNVs are expressed (gene $FDR < 0.1$).

132

133 **Top genes with pSNVs are involved in RNA splicing, virus infection and host immune**
134 **response**

135 We studied the genes most affected by pSNVs in SARS-CoV-2-associated phosphosites by
136 evaluating the null hypothesis that none of the pSNVs per gene caused sequence motif alterations
137 in the resulting protein. We computed significance scores that assigned more weight to genes
138 with multiple pSNVs affecting kinase binding motifs or replacing central phospho-residues, as
139 log-sums of complementary posterior probabilities. Finally, the significance scores were adjusted
140 for multiple testing using the false discovery rate (FDR) and filtered to select genes with
141 significant pSNVs.

142 Gene prioritization revealed 77 genes with pSNVs predicted to alter protein phosphorylation
143 through central residue or sequence motif alterations ($FDR < 0.1$), collectively including nearly
144 one third of pSNVs (**Figure 2A, Supplementary Table 3**). RNA splicing and anti-viral defense
145 responses were prominent among the functions of top genes. The two most significant genes,
146 *SRRM2* and *BUD13*, encode subunits of the spliceosome complex and include 83 and 11 pSNVs,
147 respectively ($FDR < 10^{-23}$). *SRRM2* is involved in human immunodeficiency virus (HIV)
148 pathogenesis via alternative splicing²⁵. *BUD13* regulates the antiviral transcription factor *IRF7*
149 and interferon I response upon RNA-virus infection^{26,27}. Interferon signaling triggers the first
150 response of host cells to viral infection and activates the immune system of adjacent cells to
151 suppress viral replication. The host RNA splicing machinery is targeted by SARS-CoV-2
152 proteins to disrupt splicing and impair host gene translation^{28,29}. Another related candidate gene
153 *BCLAF1* encodes a pro-apoptotic transcription and splicing factor and includes 18 pSNVs (FDR
154 = 4.2×10^{-6}). Apoptosis is triggered in infected host cells as a final form of cellular defense,

155 while activation of anti-apoptotic pathways is a strategy to maximize viral replication ³⁰.
156 BCLAF1 is also involved in type I interferon signaling and regulates antiviral gene expression
157 upon virus infection ³¹. The essential role of BCLAF1 in lung development ³² and expression in
158 lung cells suggests its activity in airway tract tissues affected by SARS-CoV-2 infection. Thus,
159 some top genes with pSNVs are involved in core host cellular processes of virus infection and
160 host immune response.

161 We highlighted the transcriptional repressor *TRIM28* with five pSNVs as a gene of interest (*FDR*
162 = 0.042) (**Figure 2B-C**). *TRIM28* increases interferon beta and pro-inflammatory cytokine
163 production in response to avian virus infection in lung epithelial cells through phosphorylation of
164 the amino acid residue S473 ³³. This phosphosite is also modified upon SARS-CoV-2 infection
165 in Vero6 cells ⁹ . One direct pSNV in *TRIM28* affects the S473 phosphosite: the amino acid
166 substitution S473L removes the central phospho-residue and causes loss of phosphorylation. The
167 site S473 is phosphorylated by checkpoint kinases (CHEK1/2) and the MAPKAP2 kinase in
168 DNA damage response and interferon activation ^{33,34}. In agreement with those experimentally
169 verified kinase binding sites, MIMP analysis predicts that the sequence motifs of the same
170 kinases are disrupted by the pSNV, including motifs of CHEK1, CHEK2 and MAPKAP2
171 (MIMP posterior *prob* ≥ 0.99) (**Figure 2C**). Another pSNV in *TRIM28*, R472C, occurs upstream
172 of the S473 phosphosite; however, unlike S473, this substitution does not cause significant motif
173 alterations. A recent study showed that knockdown of *TRIM28* activates ACE2 and leads to
174 increased host cell entry of SARS-CoV-2 ³⁵. *TRIM28* expression also correlates with interferon
175 levels, being lower in children with severe COVID-19 disease and multisystem inflammatory
176 syndrome (MIS-C) compared to uninfected children and those with mild disease ³⁶. The pSNVs
177 in the phosphosites of *TRIM28* may alter the signaling of this protein and suppress the activation
178 of immune response. Further study of this protein and its pSNVs may offer mechanistic insights
179 to its role in SARS-CoV-2 infection and COVID-19.

180

181 **Genes with frequent pSNVs are broadly expressed and enriched in host processes of virus** 182 **life cycle**

183 To interpret the human genetic variation of signaling networks responding to SARS-CoV-2
184 infection, we asked if the genes with frequent pSNVs converged to biological processes,

185 molecular pathways, and protein complexes. Pathway enrichment analysis highlighted 50 gene
186 sets with frequent pSNVs and also captured additional genes with pSNVs that remained sub-
187 significant in the gene-focused analysis ($FDR < 0.1$ from ActivePathways³⁷) (**Figure 2D,**
188 **Supplementary Table 4**). Post-transcriptional regulation and RNA splicing represented the
189 largest group of pathways and included the spliceosome complex with 120 pSNVs in seven
190 proteins (CORUM:351, $FDR = 0.020$), extending our observations from the top gene list.
191 Second, gene sets involved in basement membrane, cytoskeleton and microtubule organization
192 were enriched, perhaps reflecting the structural changes in cellular organization upon virus
193 infection⁹. The third major group involved Ras and Rho GTPase signaling proteins that respond
194 to extracellular stimuli to activate diverse cellular pathways such as proliferation, migration,
195 apoptosis, and cell adhesion. The Reactome pathway *COPI-dependent Golgi-to-ER retrograde*
196 *traffic* ($FDR = 0.031$; R-HSA-6811434) with 19 pSNVs in eight genes included the kinesins
197 *KIF1A* and *KIF3B* involved in intracellular transport, and the GTPase signaling protein
198 *RACGAP1*. SARS-CoV-2 uses the host translational machinery in the endoplasmic reticulum
199 (ER) for replication, resulting in increased ER stress and activation of unfolded protein response
200^{30,38}. Gene sets related to chromatin organization, DNA repair, cell adhesion, and mitotic cell
201 cycle processes were also enriched. Thus, pSNVs affect multifunctional proteins involved in core
202 host processes of the virus life cycle.

203 Genes with frequent pSNVs genes were expressed in diverse human tissues and cell types
204 according to functional enrichment analysis tissue-specific gene expression signatures in Human
205 Protein Atlas³⁹ (ActivePathways $FDR < 0.1$) (**Figure 2E, Supplementary Table 4**). The
206 prioritized genes were often expressed in lung pneumocytes and their squamous epithelial
207 precursors, confirming their relevance to respiratory tissues affected by SARS-CoV-2 infection.
208 The genes were also often expressed in lymph nodes and spleen, which are damaged by SARS-
209 CoV-2 infection⁴⁰. Brain and nervous system tissues, including cerebral cortex, hippocampus,
210 and peripheral nerves were also identified among enriched gene expression signatures, in line
211 with evidence of broad cellular perturbations of brain tissues in severe COVID-19⁴¹. Gene
212 expression signatures of skin, kidney, pancreas, colon, liver, and female reproductive tissues
213 were also identified. Tissue-specific enrichments were partially driven by genes with high
214 expression in most tissues, such as those involved in RNA splicing (*SRRM2*, *BCLAF1*),
215 interferon regulation (*TRIM28*), cell cycle (*NUMA1*, *SUN2*) and DNA damage response (*MSH6*).

216 SARS-CoV-2 affects diverse human tissues directly through the ACE2 receptor and indirectly
 217 through inflammation and immune response dysregulation, causing broad organ damage in
 218 severe disease⁴². Pathway annotations and expression patterns of genes with pSNVs suggest
 219 these multifunctional genes and their genetic variants may contribute to SARS-CoV-2 infection
 220 and COVID-19 through various pathways and mechanisms.

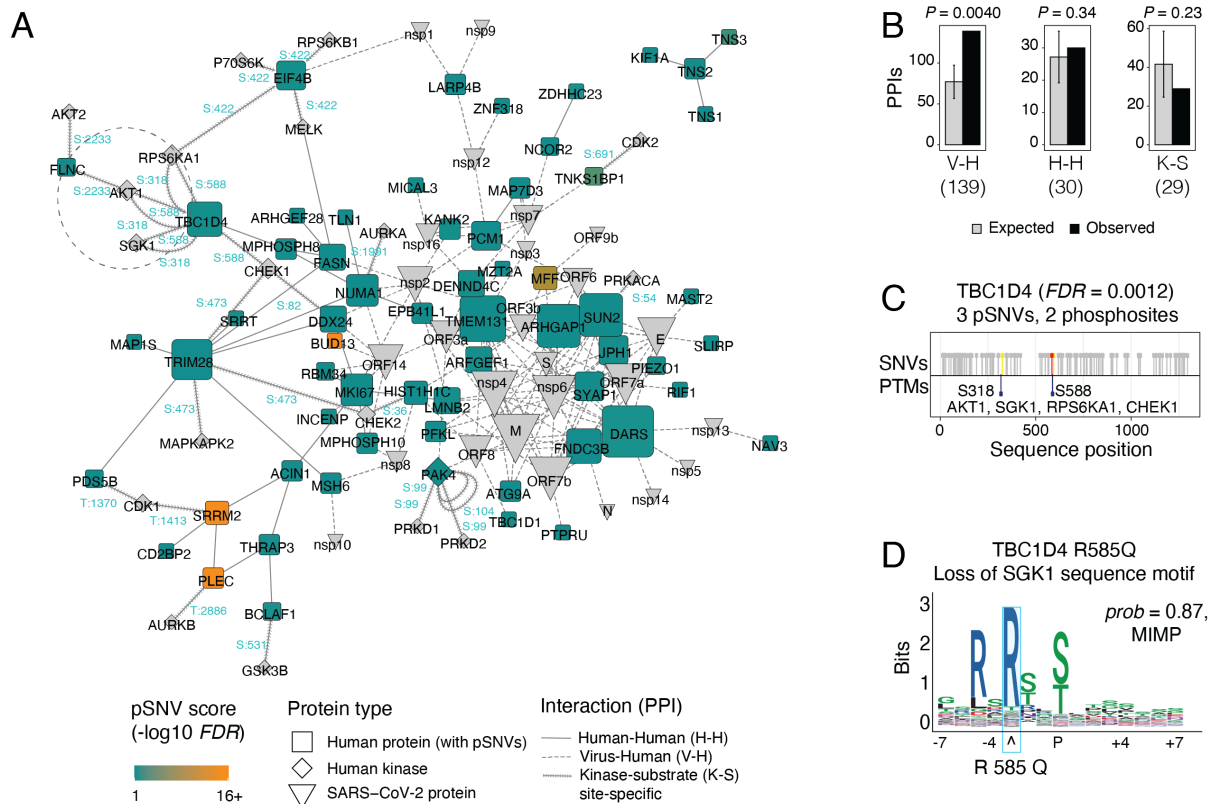


Figure 3. Protein-protein interactions (PPI) of genes with pSNVs in SARS-CoV-2 associated phosphosites.

A. The PPI network includes three classes of proteins: (a) human proteins encoded by top genes with pSNVs ($FDR < 0.1$; blue and orange squares), (b) proteins encoded by the SARS-CoV-2 virus (grey wedges), and (c) human kinases (grey diamonds) known to bind SARS-CoV-2-associated phosphosites with pSNVs. Three types of interactions are shown: PPIs of the top human proteins (H-H; solid line), PPIs of the top human proteins and SARS-CoV-2 proteins (V-H; dashed line), and site-specific kinase-substrate phosphorylation events at sites with pSNVs (K-S; arrow line). The phosphosites with pSNVs bound by specific kinases are shown in blue (e.g., S99). Node size corresponds to the number PPIs per protein (i.e., node degree). **B.** Enrichment analysis of PPIs. The PPI network is significantly enriched in host-virus interactions of top human proteins and SARS-CoV-2 proteins (i.e., V-H interactions; left). Human-human PPIs (H-H) and site-specific kinase-substrate interactions (K-S) are not enriched. Expected PPI counts were derived from control protein sets drawn randomly from the human

phospho-proteome. Empirical P-values are shown. **C.** Example of kinase-substrate interactions altered by pSNVs in phosphosites of SARS-CoV-2 infection (dotted circle in panel A). The GTPase signaling protein TBC1D4 involved in glucose homeostasis and transport, diabetes, and obesity, includes three pSNVs in the phosphosites S318 and S588 phosphorylated by the kinases AKT1, CHEK1, RPS6KA1 and SGK1. **D.** The pSNV R585Q in TBC1D4 substitutes R585 adjacent to the phospho-residue and disrupts the sequence motif of SGK1, potentially disrupting glucose homeostasis regulated by TBC1D4.

221

222 **Protein-protein interaction networks with frequent SARS-CoV-2-associated pSNVs**

223 We performed a network analysis of protein-protein interactions (PPI) to understand the
224 functional context of the top genes with pSNVs (**Figure 3A, Supplementary Table 5**). We
225 included 30 physical PPIs of the 77 top human proteins, and 139 PPIs of SARS-CoV-2 proteins
226 interacting with the top human proteins collected in the BioGRID database⁴³. We also included
227 29 high-confidence kinase-substrate interactions with specific SARS-CoV-2-associated
228 phosphosites, using data from the ActiveDriverDB database⁴⁴. Kinase-substrate interactions
229 were limited to phosphosites with at least one pSNV. To evaluate the significance of this PPI
230 network, we sampled subnetworks from the human proteome using random sets of
231 phosphoproteins as controls.

232 Human proteins with pSNVs interacted with viral proteins significantly more often than expected
233 from the human phospho-proteome (139 observed PPI vs. 77 ± 20 expected (± 1 s.d.);
234 permutation $P = 0.0040$) (**Figure 3B**). The most frequently interacting human proteins (DARS,
235 TMEM131, ARHGAP1, SUN2) and viral proteins (M, nsp4, nsp6, orf7a, orf7b) each had at least
236 10 interactions in the network. SUN2 with 9 pSNVs ($FDR = 0.021$) encodes a nuclear membrane
237 protein whose over-expression blocks HIV-1 replication and induces changes in nuclear shape⁴⁵.
238 RNA-binding proteins and spliceosome subunits such as BUD13, DDX24, and RBM34 interact
239 with orf14, a currently uncharacterized SARS-CoV-2 protein^{26,46}. DDX24 is involved in RNA
240 packaging of HIV into virions⁴⁷. RBM34 has been implicated in Middle Eastern Respiratory
241 Syndrome (MERS) infections via interactions with the viral protein nsp3.2⁴⁸. The SARS-CoV-2
242 non-structural proteins nsp4 and nsp6, frequently interacting with top pSNV-enriched host
243 proteins, localize to ER, and are involved in autophagy and viral replication^{49,50}. nsp6 suppresses
244 interferon regulatory factor 3 (IRF3) phosphorylation by binding TANK binding kinase 1

245 (TBK1)¹⁵. Thus, the human proteins with frequent pSNVs are involved in infection and
246 pathogenesis through interactions with viral proteins, suggesting that some pSNVs could impact
247 SARS-CoV-2 virus-host protein interactions and signaling pathways.

248 We examined the site-specific kinase-substrate interactions for additional details of pSNVs. The
249 enriched pathway *Translocation of SLC2A4 (GLUT4) to the plasma membrane* (Reactome: R-
250 HSA-1445148) with pSNVs in *TBC1D1*, *TBC1D4*, *RALGAPA2*, *KIF3B* and others ($FDR = 0.02$;
251 ActivePathways) identified in our pathway analysis is also apparent in the PPI network. *TBC1D4*
252 (*i.e.*, AS160), encodes a GTPase-activating protein that regulates glucose homeostasis through
253 insulin-dependent trafficking of GLUT4⁵¹. Protein-coding SNVs in *TBC1D4* and *TBC1D1* have
254 been associated with insulin resistance, obesity, and type II diabetes^{52,53}. Type II diabetes and
255 obesity are SARS-CoV-2 comorbidities and may implicate certain individuals and populations as
256 more susceptible to severe COVID-19⁴⁵. In our data, *TBC1D4* includes three pSNVs that affect
257 the SARS-CoV-2-associated phosphosites S318 and S588 ($FDR = 0.0011$). The sites are
258 phosphorylated by AKT1, SGK1, RPS6KA1, and CHEK1 kinases based on earlier experimental
259 studies⁵⁴⁻⁵⁶ (**Figure 3C**). The pSNV R585Q disrupts sequence motifs in the *TBC1D4*
260 phosphosite S588, including motifs bound by the same kinases (CHEK1, SGK1, RPS6KA1)
261 (**Figure 3D**). *TBC1D4* phosphorylation at S588 is known to regulate GLUT4 translocation and
262 subsequent activity in adipocytes⁵¹. Thus, the predicted motif disruption via pSNVs matches the
263 kinases binding the phosphosite. The related protein *TBC1D1* has nine pSNVs in the
264 phosphosites S614, S627, and S660, including four pSNVs potentially disrupting
265 phosphorylation ($FDR = 3.8 \times 10^{-5}$); however, the kinases binding these phosphosites are
266 unknown. This example suggests that some pSNVs provide mechanistic hypotheses for studying
267 SARS-CoV-2 comorbidities.

268

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

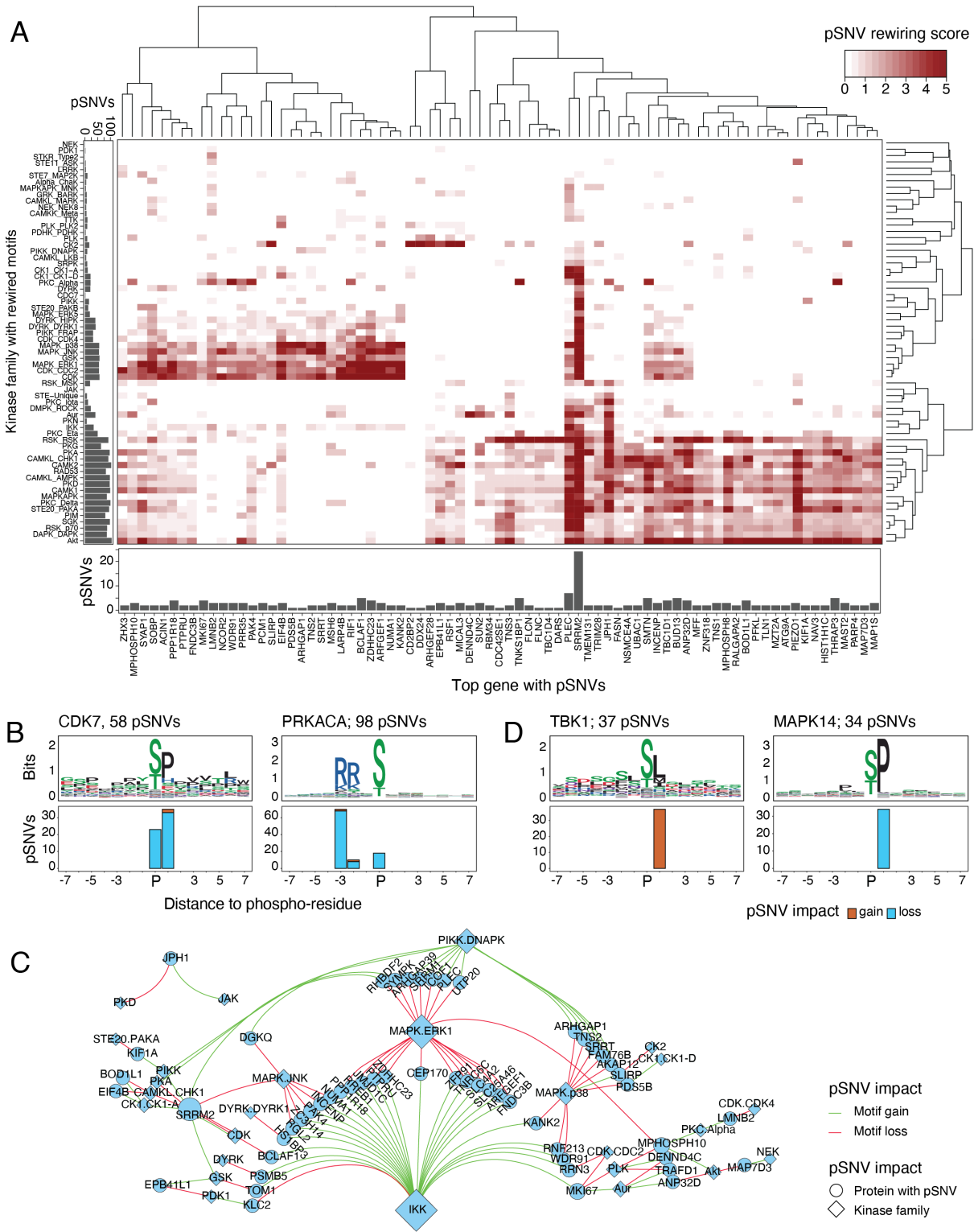


Figure 4. Kinases affected by pSNV-driven rewiring of sequence motifs in SARS-CoV-2 associated phosphosites. A. Heatmap shows the hierarchical clustering of top genes with pSNVs (X-axis), and kinase

families affected by sequence motif rewiring by pSNVs (Y-axis). Genes and kinases are clustered by rewiring scores that reflect the pSNVs and their posterior probabilities of motif rewiring or central residue removal. Two clusters of kinase families are subject to motif rewiring: MAPK and CDK (middle-left) and PRKA, CAMK, CHEK, AKT, and RPS (bottom-right). Bar plots show pSNVs in sequence motifs (left) and top genes (bottom). Since motifs of different kinase families are often similar and accounted for, pSNV counts per kinase family are redundant. Scores are capped at 5 for visualization. **B.** Examples of sequence motifs rewired by pSNVs for the Cyclin Dependent Kinase 7 (CDK7) (left), Protein Kinase CAMP-Activated Catalytic Subunit Alpha (PRKACA) (middle), Calcium/Calmodulin Dependent Protein Kinase II Alpha (CAMK2A) (right). Sequence logo of the position weight matrix (PWM) bound by the kinase (top) and bar plot with pSNVs altering the motif (bottom) are shown. **C.** Network of pSNVs causing motif switches (*i.e.*, combined gain of one motif and loss of another motif). Nodes show kinase families (diamonds) and top genes (circles), and edges show pSNV-driven motif losses and gains (red and green, respectively). Node size indicates the number of gained and lost motifs (*i.e.*, node degree). For each pSNV, only the kinase family with the highest motif rewiring probability is shown. **D.** Sequence motifs bound by the TANK binding kinase 1 (TBK1) of the IKK family are consistently gained through pSNVs (left) while motifs bound by CDK and MAPK kinase families are lost through the same pSNVs. Bar plots show the pSNVs causing gains of TBK1 motifs (left) with corresponding losses of MAPK14 motifs (right). Similar motifs of other kinase families replaced by TBK1 motifs are not shown. pSNVs in top genes are included in panels A and B, and pSNVs in all genes are included in panels C and D.

269

270 **Motif-switching pSNVs exchange CDK and MAPK kinase targets in favor of TBK1**

271 To characterize the effect of pSNVs on rewiring SARS-CoV-2 phosphorylation networks, we
272 studied the kinase families whose sequence motifs were most affected by pSNVs in top genes.

273 We computed significance scores for kinase-substrate pairs by accounting for all high-
274 confidence motif predictions, giving a higher priority to kinase-substrate pairs where the same
275 type of sequence motif was affected by multiple pSNVs of the protein. This strategy included
276 similar motifs of evolutionarily related kinase families and led to partially redundant scoring of
277 kinase motifs. Collectively, we included 217 pSNVs with 3360 motif-rewiring predictions
278 covering 64 kinase families.

279 Hierarchical clustering of motif-rewiring pSNVs revealed two major clusters of kinases and
280 substrate proteins (**Figure 4A**). The first cluster was characterized by pSNVs causing losses of
281 binding motifs of cyclin dependent kinases (CDKs) and mitogen associated kinases (MAPK;
282 p38, Jnk, Erk) in genes encoding the proliferation marker MKI67, DNA damage response protein
283 MSH6, spliceosome subunit SRRT, interferon regulator BCLAF1, the translation initiation factor

284 EIF4B, and others. For instance, sequence binding motifs of CDK7 were lost through 56 pSNVs
285 and gained through two pSNVs (**Figure 4B**). SARS-CoV-2 infection is known to promote
286 activation of host MAPK p38 kinases and inhibition of CDKs ⁹. The second, larger cluster
287 involved motif rewiring of Protein Kinase A, CAMK, checkpoint, and PI3K/Akt kinase. The
288 PI3K/Akt pathway controls cell proliferation, survival, and suppression of apoptosis, and
289 responds to extracellular signals of cytokines and growth factors via the mTOR pathways.
290 Checkpoint kinases respond to DNA damage by blocking cell proliferation and inducing cell
291 death. Among the top proteins, motifs bound by Protein Kinase A (PRKACA) and CAMK2A
292 were affected by 98 pSNVs, most of which led to motif losses or replacements of central
293 phospho-residues (**Figure 4B**). Motif-rewiring pSNVs in this cluster occur in transcriptional and
294 post-transcriptional regulators such as PLEC, BUD13, THRAP3, and TRIM28, among others.
295 The spliceosome subunit SRRM2, with the largest number of pSNVs in our analysis, included
296 motif-rewiring pSNVs affecting kinases of both clusters. Therefore, pSNVs may reconfigure
297 kinase signaling networks of defense processes and host components of the viral life cycle.

298 We focused on pSNVs that disrupted one type of sequence motif and induced another type of
299 motif at the same mutated phosphosite, potentially causing switches in kinase binding. This
300 revealed a network of 54 proteins with pSNV-induced binding switches involving 26 kinase
301 families, when selecting the most confident kinase family for every motif-rewiring prediction
302 (**Figure 4C, Supplementary Table 6**). Motif switches were caused by 60 pSNVs. Strikingly, the
303 largest subnetwork of motif switches involved motif gains of the Tank Binding Kinase 1 (TBK1)
304 of the IKB kinase (IKK) family that replaced motifs of CDKs and MAPKs (**Figure 4D**). We
305 found 37 pSNVs with this motif-switching impact. The top proteins with motif-switching pSNVs
306 included the transcriptional regulators BCLAF1 and RREB1, the WDR91 protein involved in
307 host cell entry of SARS-CoV-2 ⁵⁷, the nuclear matrix protein NUMA, and the proliferation
308 marker MKI67. TBK1 controls immune monitoring pathways and phosphorylates interferon
309 regulatory factors IRF-3 and IRF-7 to trigger the host antiviral response that is suppressed by
310 SARS-CoV-2 proteins ^{10,11,14,15}. Thus, in some individuals and populations, motif-rewiring
311 pSNVs induce structural changes in kinase signaling networks and cross-talk of mitogenic and
312 immune response pathways. These network-rewiring pSNVs potentially reflect the inter-
313 individual heterogeneity and evolutionary pressures of signaling networks responding to viral
314 infections.

315

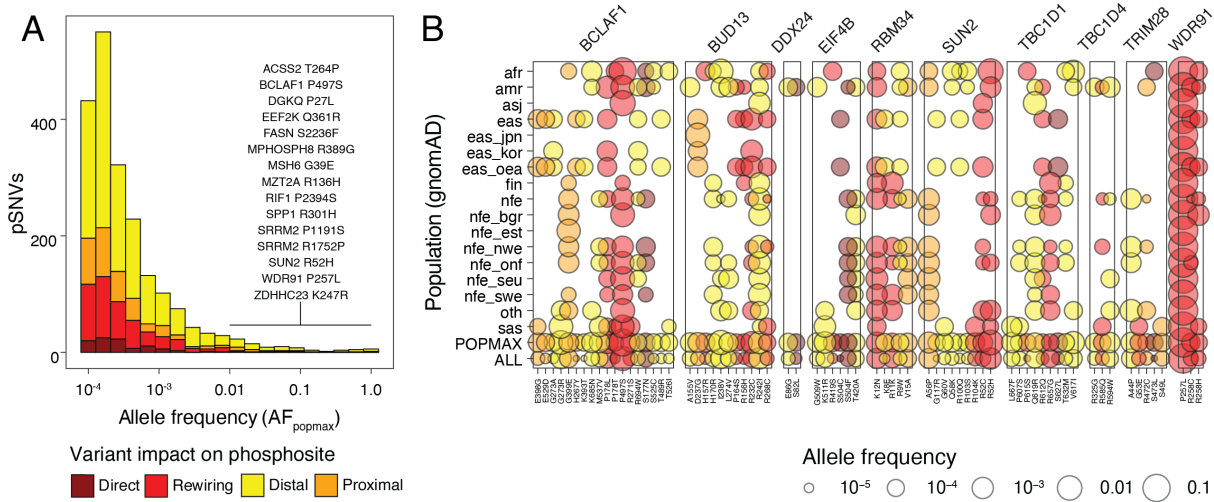


Figure 5. Allele frequencies of SARS-CoV-2-associated pSNVs in the human population. A. Histogram of allele frequencies of pSNVs in the gnomAD dataset (AF_{popmax}) grouped by functional impact of pSNVs. AF_{popmax} corresponds to the population group where the allele frequency was the highest. More-frequent and impactful pSNVs are listed, with AF_{popmax} of least 1% and a direct or motif-rewiring impact. **B.** Allele frequencies of pSNVs vary among human populations in gnomAD. Dot plot shows a subset of top genes and pSNVs discussed in the study (X-axis) with their allele frequencies in populations, the maximum population-based allele frequency (AF_{popmax}) and the total allele frequency in gnomAD (AF_{all}) (Y-axis).

316

317 Population genetic variation and disease associations of pSNVs

318 We studied the allele frequency of pSNVs to evaluate the extent of predicted changes in
 319 infection-responsive signaling networks in the 16 human populations in the gnomAD dataset
 320 (**Figure 5A**). Most pSNVs were relatively infrequent, with a median allele frequency of four per
 321 hundred thousand individuals (AF_{all} = 3.8 x 10⁻⁵) while the median frequency in the most
 322 representative population was an order of magnitude higher (AF_{popmax} = 2.1 x 10⁻⁴), suggesting
 323 the value of a deeper population-based analysis of certain pSNVs. Some pSNVs (86 or 4.2%)
 324 were relatively common and occurred at above 1% frequency in at least one population,
 325 including 15 pSNVs involved in sequence motif rewiring or loss of phospho-residues. For
 326 example, the pSNV P257L in WDR91 commonly observed in the human population (AF_{all} =
 327 0.72) causes switches of sequence motifs at the phosphosite S257, by inducing motifs of the

328 TBK1 kinase and disrupting motifs of CDK and MAPK kinases. *WDR91* encodes a WD repeat-
329 containing protein that was identified in a CRISPR screen as a host regulator of endosomal entry
330 of SARS-CoV-2⁵⁷. Frequent motif-rewiring pSNVs were found in other top proteins, such as the
331 spliceosome subunit SRRM2, the pro-apoptotic splicing and transcription factor BCLAF1, and
332 the HIV-related nuclear membrane protein SUN2. pSNV frequency varied among populations in
333 gnomAD (**Figure 5B**). For example, the pSNV S473L in TRIM28 replacing the phosphoresidue
334 S473 involved in interferon regulation was exclusively found in the African population ($AF_{\text{afr}} =$
335 1.2×10^{-4}). The adjacent pSNV R472C detected in Latino, North-Western European, and other
336 populations ($10^{-5} < AF < 10^{-4}$) may also contribute to altered signaling, however no motif
337 disruptions were identified for this pSNV. As another example, the pSNV R585Q disrupting
338 sequence motifs in the TBC1D4 phosphosite S588 involved in glucose transport was detected in
339 Latino, South Asian and North-Western European populations ($AF_{\text{sas}} = 1.3 \times 10^{-4}$ and $AF_{\text{amr}} =$
340 2.9×10^{-5} and $AF_{\text{nfe_nwe}} = 4.8 \times 10^{-5}$) while additional pSNVs in TBC1D1 and TBC1D4 found in
341 other populations may also contribute to the rewiring of SARS-CoV-2 associated signaling
342 networks.

343 We found 86 pSNVs associated with human disease as variants of unknown significance (VUS)
344 in the ClinVar database⁵⁸ (**Supplementary Table 7**). pSNV were associated with cancer
345 predisposition (NF1, PMS2, FLCN, MSH6, APC), cardiovascular phenotypes (BMPR2, MYLK,
346 VCL, LMNA), muscular dystrophy (SUN2, FLNC), natural killer cell and glucocorticoid
347 deficiency (MCM4), autoimmune interstitial lung, joint, kidney disease (COPA), and renal cysts
348 and diabetes syndrome (HNF1B). Motif-rewiring effects of disease-associated pSNVs suggest
349 that SARS-CoV-2 and COVID-19 may hijack signal transduction networks involved in human
350 disease. However, since the disease-associated pSNVs are annotated as variants of unknown
351 significance, we have currently no evidence of their involvement in pathogenesis.

352 We asked if the pSNVs could be associated directly with risk of COVID-19, by analyzing the
353 genome-wide association study (GWAS) of the COVID-19 Host Genetics Initiative that explored
354 the genetic variation of nearly 50,000 COVID-19 patients using genotyping and imputation⁶.
355 Only a minority of pSNVs we identified were reported in the GWAS (66 or 4%). This low
356 coverage is likely explained by the overall low allele frequency of most pSNVs that restricts
357 their imputation and risk assessment in patient cohorts. We found a few pSNVs with nominal

358 statistical associations with COVID-19 risk, however, the associations were not significant after
359 multiple testing correction (all $FDR > 0.15$).

360 Collectively, these data suggest that pSNVs may mediate topological variation in signaling
361 networks that respond to SARS-CoV-2 infection. Since most pSNVs occur infrequently in the
362 human population, associating these with clinical risk using current datasets is challenging.
363 Focused genetic association studies and mechanistic experiments are needed to confirm and
364 characterize these predictions, and further evaluate their clinical potential.

365 **DISCUSSION**

366 We highlighted human genetic polymorphisms that may cause rewiring of phospho-signaling
367 networks responding to SARS-CoV-2 infection, using a machine-learning approach that
368 identified impactful mutations in protein sequence motifs. The proteome-wide enrichment of
369 motif-rewiring pSNVs suggests evolutionary plasticity of the human signaling network
370 responding to virus infection. This evolution of host defense strategies is exemplified by the
371 number of motif-switching pSNVs that induce novel substrate sites of the TANK Binding Kinase
372 1 in favor of motifs recognized by MAPK and CDK kinases. Motif-rewiring pSNVs were found
373 in core host processes of the viral life cycle such as RNA splicing and cell cycle, innate immune
374 responses such as interferon activation, Ras/GTPase signaling proteins, and in other human
375 genes involved in virus infections. Genes with pSNVs were often expressed in diverse human
376 tissues, reflecting the broad organotropism of SARS-CoV-2 and potential contribution of pSNVs
377 to multi-organ failure in severe COVID-19, as well as diverse long-term symptoms. Disease
378 genes and variants of unknown significance found among pSNVs suggest that infections may
379 hijack disease pathways, and the pSNVs in the pathways may contribute to symptomatic
380 diversity and comorbidities of COVID-19.

381 Our study has important limitations. We currently have no statistical or clinical evidence of
382 pSNVs associating with SARS-CoV-2 infection, COVID-19 risk or comorbidities, as most
383 pSNVs are infrequent in human populations and remain unmapped in genome-wide association
384 studies. The phosphosites we studied reflect an early post-infection timepoint in cell culture,
385 limiting our analysis of signaling pathways activated further downstream of SARS-CoV-2
386 infection in human tissues and the immune system. On the one hand, our analysis may
387 overestimate the extent of network rewiring because our sequence-based pSNV impact analysis
388 does not account for co-expression and localization of kinases and substrates. On the other hand,
389 the number of motif-rewiring pSNVs may be underestimated, because many motifs bound by
390 kinases and other phospho-enzymes remain unknown. The known landscape of functional
391 protein-coding SNVs may be extended by analyzing other types of PTMs.

392 Our integrative proteogenomic analysis of pSNVs affecting signaling networks of SARS-CoV-2
393 enables further studies of viral infection, disease mechanisms and potential translation. Analysis
394 of whole-genome sequencing datasets with matched clinical profiles of COVID-19 information

395 is needed to associate our functional predictions of infrequent pSNVs with patient risk and co-
396 morbidities, and to enable the development of prognostic and predictive biomarkers. The
397 candidate genes, pathways, and kinases identified in this study enable functional genomics
398 screens and phenotypic experiments that may lead to further mechanistic understanding of
399 SARS-CoV-2 infection and COVID-19. The biological processes and kinase signaling networks
400 we identified involve known drug targets, some of which have already been proposed in recent
401 studies. Thus, our findings have the potential to contribute to genome-guided risk predictions and
402 development of therapies for the current and future pandemics.

403 METHODS

404 Host phosphorylation sites. Protein phosphosites significantly differentially phosphorylated in
405 SARS-CoV-2 infection were retrieved from the phospho-proteomics study by Bouhaddou *et al*⁹.
406 We mapped the phosphosites to canonical protein isoforms (hg19) using the ActiveDriverDB⁴⁴
407 database. SARS-CoV-2 associated phosphosites detected at the 24h timepoint were filtered for
408 significance ($FDR < 0.05$ in infected cells; $FDR > 0.05$ in controls). Sites corresponding to non-
409 phosphorylated residues in human proteins were removed (*i.e.*, residues other than S, T, or Y).
410 To compare observed and expected numbers of pSNVs, all experimentally detected human
411 phosphosites were retrieved from ActiveDriverDB based on proteomics databases
412 (PhosphoSitePlus⁵⁹, UniProt⁶⁰, Phospho.ELM⁶¹, HPRD⁶²). SARS-CoV-2 associated
413 phosphosites were excluded from these controls. Known interactions of kinases and specific
414 phosphosites were also retrieved from ActiveDriverDB, based on multiple databases⁵⁹⁻⁶².

415 Human genome variation. Genome variation maps of human populations were retrieved from the
416 gnomAD exome sequencing project of 124,748 individuals (version 2.1.1; hg19)²³. Single
417 nucleotide variants (SNVs) were first filtered by variant quality (*i.e.*, by selecting variants where
418 filter = PASS) and maximum allele frequency in the gnomAD populations ($AF_{popmax} \geq 10^{-4}$). The
419 ANNOVAR software⁶³ was used to annotate the SNVs in protein-coding genes. Missense SNVs
420 in canonical protein isoforms were selected to match the protein phosphosites. We excluded
421 small insertions-deletions (indels), splicing, frameshift, and stop codon mutations, and missense
422 SNVs with mismatching reference amino acid residues.

423 pSNV impact prediction. Phosphorylation associated SNVs (pSNVs) occurred in SARS-CoV-2-
424 associated phosphosites in flanking sequences of ± 7 residues. Four mutually exclusive categories
425 of variant impact were assigned in order of priority: direct (substitution of phosphorylated
426 residues S, T, Y), motif-rewiring (gain or loss of motif, based on MIMP²⁴ analysis), and
427 proximal and distal (within 1-2 and 3-7 residues from phosphosite, respectively, with no motif-
428 rewiring prediction). In the cases where multiple adjacent phosphosites were found to match a
429 pSNV, the highest-impact phosphosite was used. Motif-rewiring predictions were derived using
430 the MIMP method²⁴, with the database of high-confidence kinase position weight matrices
431 (PWMs) (by setting the parameter model.data = hconf), a posterior probability cutoff ($prob >$

432 0.5), and inclusion of central residues for rewiring predictions (include.cent = TRUE). To score
433 each pSNV, the kinase motif (*i.e.*, the PWM) with the largest posterior probability was selected.
434 To prioritize pSNVs causing motif switches, both posterior probabilities (*i.e.*, the top motif gain
435 and loss) were selected and aggregated as $prob_{\text{switch}} = 1 - (1 - prob_{\text{gain}})(1 - prob_{\text{loss}})$,
436 corresponding to the combined probability of the pSNV causing either a gain or a loss of a motif.

437 Evaluating the enrichment of motif-rewiring pSNVs. Randomly sampled phosphosites were used
438 to evaluate the expected numbers of pSNVs with functional impact predictions (*i.e.*, direct,
439 motif-rewiring, proximal, and distal pSNVs). We randomly sampled 10,000 sets of phosphosites
440 from the human phospho-proteome using datasets from ActiveDriverDB as controls, in equal
441 numbers to SARS-CoV-2 associated phosphosites. Control sites were annotated using gnomAD
442 SNVs and MIMP analysis, similarly to SARS-CoV-2-associated phosphosites, to derive the
443 expected distributions of annotated pSNVs and empirical *P*-values comparing observed and
444 expected values were reported. Expected counts were shown as mean values with ± 1 standard
445 deviation for confidence intervals.

446 Gene prioritization. Genes were scored based on the aggregated impact of pSNVs in
447 phosphosites. Each pSNV was assigned either the posterior probability of motif rewiring
448 ($prob_{\text{SNV}} = prob_{\text{rewiring}}$ from MIMP; see above), or alternatively, a conservative posterior
449 probability of motif rewiring ($prob_{\text{SNV}} = 0.05$) if it was annotated as distal or proximal to the
450 phosphosite with no motif-rewiring predictions from MIMP. To derive gene-level significance
451 scores (*P*-values), posterior probabilities of all pSNVs in each gene were aggregated as $P_{\text{gene}} = 1$
452 $- \prod_{\text{SNV of gene}} (1 - prob_{\text{SNV}})$, reflecting the null hypothesis that none of the pSNVs in the protein
453 encoded by the gene independently altered kinase binding motifs. The resulting gene-level *P*-
454 values were corrected for multiple testing with the Benjamini-Hochberg false discovery rate
455 (*FDR*) method and genes with significant *FDR* values were selected ($FDR < 0.1$). *FDR* values
456 were capped at 10^{-16} for visualization.

457 Functional enrichment analysis. Pathway enrichment analysis was conducted using the
458 ActivePathways method³⁷ and included gene sets corresponding to biological processes of Gene
459 Ontology⁶⁴, molecular pathways of Reactome⁶⁵, protein complexes of CORUM⁶⁶, and genes
460 expressed in human tissues of the Human Protein Atlas³⁹. Gene sets were retrieved from the

461 g:Profiler web server ⁶⁷ (Mar 25th, 2021). Tissue-specific gene expression signatures and
462 functional gene sets were analyzed separately. As the statistical background, only experimentally
463 determined human phosphoproteins in ActiveDriverDB were used for increased stringency.
464 Significantly enriched pathways were detected using default parameters of ActivePathways and
465 significant gene sets were selected (ActivePathways $FDR < 0.1$). Functional gene sets were
466 visualized as an enrichment map ⁶⁸, functional themes were annotated manually, and the highest-
467 ranking genes with pSNVs were shown (gene $FDR < 0.25$). Genes with pSNVs highly expressed
468 in human tissues were visualized separately. FDR values were capped at 10^{-16} for visualization.

469 Analysis of protein-protein interaction (PPI) networks. Three types of PPIs were included to
470 analyze the network of top genes with pSNVs ($FDR < 0.1$). First, we included physical human-
471 human PPIs where both interacting host proteins were part of the top gene list. Second, we
472 included physical human-virus PPIs where one interacting host protein was part of the top gene
473 list, and the other protein was encoded by the SARS-CoV-2 virus. Third, we included
474 experimentally verified site-specific kinase-substrate interactions where the substrate host
475 protein was part of the top gene list, the host kinase protein was verified to bind a phosphosite of
476 the substrate protein based on previous experimental studies, and the phosphosite had at least one
477 pSNV. Human-human and human-virus PPIs were retrieved from the BioGRID database ⁴³ (V
478 4.4.199, downloaded on June 29th, 2021). PPIs of the third type were retrieved from
479 ActiveDriverDB. The PPI network was visualized using the Cytoscape software ⁶⁹. Empirical P -
480 values were used to evaluate the enrichment of PPIs among the top genes with pSNVs. Control
481 sets of proteins were randomly drawn from the human phospho-proteome in equal numbers to
482 top genes with pSNVs and control PPI networks were constructed. PPI counts in these networks
483 provided the expected distributions and empirical P -values for human-human PPIs, human-virus
484 PPIs, and human kinase-substrate PPIs with pSNVs. Expected counts were shown as mean
485 values with ± 1 standard deviation for confidence intervals.

486 Analysis of population frequency and disease annotations of pSNVs. Allele frequencies of
487 pSNVs in human populations were reported as shown in the gnomAD dataset. We also studied
488 disease associations of pSNVs recorded in the ClinVar database ⁵⁸ that were previously mapped
489 to canonical protein isoforms in ActiveDriverDB. Disease annotations of pSNVs were then
490 reviewed manually using the ClinVar website (data retrieved on Oct 6th, 2021). Based on this

491 review, lower-confidence pSNV annotations with no ClinVar star ratings, and annotations with
492 consensus rating of *benign* were filtered.

493 **SUPPLEMENTARY INFORMATION**

494 **Supplementary Table 1. Phosphorylation associated SNVs (pSNVs) of SARS-CoV-2**

495 **infection.** The table shows all the pSNVs in the phosphosites affected in SARS-CoV-2 infection
496 in the gnomAD dataset detected in at least one human population ($AF_{popmax} \geq 10^{-4}$). The table
497 includes allele frequencies of pSNVs in the 16 populations in gnomAD (columns *AF_all*,
498 *AF_popmax*, etc.), the kinases known to target the site (column *ptm_kinase*), the probability of
499 the pSNV disrupting the phosphosite based on MIMP analysis (column *top_MIMP_score*; the
500 highest-ranking kinase motif), the predicted impact of the pSNV on the phosphorylation (column
501 *pSNV_impact*), and whether the pSNV affects multiple phosphosites (column *multi_PTM*). In
502 case of multiple sites, the phosphosite with highest MIMP probability is used for scoring (see
503 Supplementary Table 2).

504 **Supplementary Table 2. Motif-rewiring predictions of pSNVs.** Predictions of the impact of
505 pSNVs on sequence motifs bound by kinases, derived from the MIMP method. Provided are the
506 amino acids in the phosphosites of reference and mutated sequences, the corresponding MIMP
507 scores of the reference and mutated sequences of being bound by kinases, the ratio of scores
508 (column *log_ratio*), and the posterior probability of motif rewiring (column *prob*). Missing
509 values (NA) are shown for pSNV scores where the impact is direct removal of a phosphoresidue.

510 **Supplementary Table 3. Top genes with pSNVs.** Complete list of genes with pSNVs ranked
511 by their *FDR* scores, corresponding to the null hypothesis that no pSNVs per gene cause a motif
512 rewiring at SARS-CoV-2-associated phosphosites. Filtered genes ($FDR < 0.1$) are used in most
513 analyses.

514 **Supplementary Table 4. Molecular pathways, processes, and tissue-specific gene expression**
515 **profiles of top genes with pSNVs.** Results of pathway enrichment analysis of top genes with
516 pSNV using ActivePathways ($FDR < 0.1$). Gene sets of biological processes of Gene Ontology,
517 molecular pathways of Reactome, protein complexes of CORUM, and tissue specific gene
518 expression signatures from the Human Protein Atlas (HPA) database are shown. For each
519 enriched gene set, the *FDR*-ranked list of genes with pSNVs in the gene set is included.

520 **Supplementary Table 5. Protein-protein interaction (PPI) network of top genes with**
521 **pSNVs.** Three types of interactions are shown (column *interaction_type*): physical interactions

522 of top genes (proteins) with pSNVs (*i.e.*, *human_human*), physical interactions of top genes with
523 pSNVs and proteins encoded by SARS-CoV-2 (*i.e.*, *human_sars*), and site-specific signaling
524 interactions of top genes with pSNVs and known kinases binding these phosphosites based on
525 information collected in ActiveDriverDB (*i.e.*, *w_kinase*). Site-specific interactions have at least
526 one pSNV in the phosphosites. The scores are $-\log_{10}$ -transformed *FDR* values of top genes.

527 **Supplementary Table 6. Motif-switching pSNVs in top genes.** The subset of pSNVs causing
528 motif switching, resulting in the joint loss of one type of motif and the gain of another motif. For
529 each pSNV, only the kinase and kinase family of the highest-scoring prediction is shown (see
530 Supplementary Table 2 for all predictions).

531 **Supplementary Table 7. Disease annotations of all pSNVs.** Disease annotations of pSNVs
532 from the ClinVar database are shown. Consensus evaluation of disease significance was based on
533 a systematic analysis of ClinVar annotations in ActiveDriverDB and a manual review of the
534 ClinVar website (column *significance_curated*). Most disease-associated pSNVs are annotated
535 as variants of unknown significance (VUS). Multiple disease annotations of the same pSNV are
536 concatenated (column *disease*).

537

538 **Acknowledgments.** This work was supported by the COVID-19 Supplements of the Project
539 Grant of Canadian Institutes of Health Research (CIHR) to J.R. and the Discovery Grant of the
540 Natural Sciences and Engineering Research Council (NSERC) to J.R., as well as the Investigator
541 Award to J.R. from the Ontario Institute for Cancer Research (OICR). A.B. was supported by an
542 Ontario Graduate Scholarship. M.K. was supported by the Scatcherd European Scholarship.
543 Funding to OICR is provided by the Government of Ontario.

544 **Author contributions.** D.P., A.B. and J.R. analyzed the data, interpreted the results, and
545 prepared the figures. M.K. preprocessed the data and led the database development. D.P., A.B.,
546 and J.R. wrote the manuscript. J.R. conceived and supervised the project. All authors reviewed
547 and edited the manuscript and approved the final version.

548 References

- 549 1 Richardson, S. *et al.* Presenting Characteristics, Comorbidities, and Outcomes Among
550 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* **323**,
551 2052-2059, doi:10.1001/jama.2020.6775 (2020).
- 552 2 Webb Hooper, M., Naples, A. M. & Perez-Stable, E. J. COVID-19 and Racial/Ethnic
553 Disparities. *JAMA* **323**, 2466-2467, doi:10.1001/jama.2020.8598 (2020).
- 554 3 Pareek, M. *et al.* Ethnicity and COVID-19: an urgent public health research priority. *The*
555 *Lancet* **395**, 1421-1422, doi:10.1016/s0140-6736(20)30922-3 (2020).
- 556 4 Nicola, M. *et al.* The socio-economic implications of the coronavirus pandemic (COVID-
557 19): A review. *Int J Surg* **78**, 185-193, doi:10.1016/j.ijisu.2020.04.018 (2020).
- 558 5 Severe Covid Gwas Group *et al.* Genomewide Association Study of Severe Covid-19
559 with Respiratory Failure. *N Engl J Med* **383**, 1522-1534, doi:10.1056/NEJMoa2020283
560 (2020).
- 561 6 Covid-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-
562 19. *Nature*, doi:10.1038/s41586-021-03767-x (2021).
- 563 7 Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature*
564 **591**, 92-98, doi:10.1038/s41586-020-03065-y (2021).
- 565 8 Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome networks and human disease.
566 *Cell* **144**, 986-998, doi:10.1016/j.cell.2011.02.016 (2011).
- 567 9 Bouhaddou, M. *et al.* The Global Phosphorylation Landscape of SARS-CoV-2 Infection.
568 *Cell* **182**, 685-712 e619, doi:10.1016/j.cell.2020.06.034 (2020).
- 569 10 Sharma, S. *et al.* Triggering the interferon antiviral response through an IKK-related
570 pathway. *Science* **300**, 1148-1151, doi:10.1126/science.1081315 (2003).
- 571 11 Balka, K. R. *et al.* TBK1 and IKKepsilon Act Redundantly to Mediate STING-Induced
572 NF-kappaB Responses in Myeloid Cells. *Cell Rep* **31**, 107492,
573 doi:10.1016/j.celrep.2020.03.056 (2020).
- 574 12 Matsuyama, T., Kubli, S. P., Yoshinaga, S. K., Pfeffer, K. & Mak, T. W. An aberrant
575 STAT pathway is central to COVID-19. *Cell Death Differ* **27**, 3209-3225,
576 doi:10.1038/s41418-020-00633-7 (2020).
- 577 13 Appelberg, S. *et al.* Dysregulation in Akt/mTOR/HIF-1 signaling identified by proteo-
578 transcriptomics of SARS-CoV-2 infected cells. *Emerg Microbes Infect* **9**, 1748-1760,
579 doi:10.1080/22221751.2020.1799723 (2020).
- 580 14 Lei, X. *et al.* Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat*
581 *Commun* **11**, 3810, doi:10.1038/s41467-020-17665-9 (2020).
- 582 15 Xia, H. *et al.* Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep* **33**, 108234,
583 doi:10.1016/j.celrep.2020.108234 (2020).
- 584 16 Hirano, T. & Murakami, M. COVID-19: A New Virus, but a Familiar Receptor and
585 Cytokine Release Syndrome. *Immunity* **52**, 731-733, doi:10.1016/j.immuni.2020.04.003
586 (2020).
- 587 17 Mishra, S., Bassi, G. & Nyomba, B. G. Inter-proteomic posttranslational modifications of
588 the SARS-CoV-2 and the host proteins A new frontier. *Exp Biol Med (Maywood)* **246**,
589 749-757, doi:10.1177/1535370220986785 (2021).
- 590 18 Li, S., Iakoucheva, L. M., Mooney, S. D. & Radivojac, P. Loss of post-translational
591 modification sites in disease. *Pac Symp Biocomput*, 337-347 (2010).

- 592 19 Reimand, J. & Bader, G. D. Systematic analysis of somatic mutations in phosphorylation
593 signaling predicts novel cancer drivers. *Molecular systems biology* **9**, 637,
594 doi:10.1038/msb.2012.68 (2013).
- 595 20 Creixell, P. *et al.* Kinome-wide decoding of network-attacking mutations rewiring cancer
596 signaling. *Cell* **163**, 202-217, doi:10.1016/j.cell.2015.08.056 (2015).
- 597 21 Huang, K. L. *et al.* Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **173**,
598 355-370 e314, doi:10.1016/j.cell.2018.03.039 (2018).
- 599 22 Reimand, J., Wagih, O. & Bader, G. D. Evolutionary constraint and disease associations
600 of post-translational modification sites in human genomes. *PLoS Genet* **11**, e1004919,
601 doi:10.1371/journal.pgen.1004919 (2015).
- 602 23 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in
603 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).
- 604 24 Wagih, O., Reimand, J. & Bader, G. D. MIMP: predicting the impact of mutations on
605 kinase-substrate phosphorylation. *Nat Methods* **12**, 531-533, doi:10.1038/nmeth.3396
606 (2015).
- 607 25 Wojcechowskyj, J. A. *et al.* Quantitative phosphoproteomics reveals extensive cellular
608 reprogramming during HIV-1 entry. *Cell Host Microbe* **13**, 613-623,
609 doi:10.1016/j.chom.2013.04.011 (2013).
- 610 26 Zhang, X. *et al.* Structure of the human activated spliceosome in three conformational
611 states. *Cell Res* **28**, 307-322, doi:10.1038/cr.2018.14 (2018).
- 612 27 Frankiw, L. *et al.* BUD13 Promotes a Type I Interferon Response by Countering Intron
613 Retention in Irf7. *Mol Cell* **73**, 803-814 e806, doi:10.1016/j.molcel.2018.11.038 (2019).
- 614 28 Banerjee, A. K. *et al.* SARS-CoV-2 Disrupts Splicing, Translation, and Protein
615 Trafficking to Suppress Host Defenses. *Cell* **183**, 1325-1339 e1321,
616 doi:10.1016/j.cell.2020.10.004 (2020).
- 617 29 Finkel, Y. *et al.* SARS-CoV-2 uses a multipronged strategy to impede host protein
618 synthesis. *Nature* **594**, 240-245, doi:10.1038/s41586-021-03610-3 (2021).
- 619 30 Ostaszewski, M. *et al.* COVID19 Disease Map, a computational knowledge repository of
620 virus-host interaction mechanisms. *Mol Syst Biol* **17**, e10387,
621 doi:10.15252/msb.202110387 (2021).
- 622 31 Qin, C. *et al.* Bclaf1 critically regulates the type I interferon response and is degraded by
623 alphaherpesvirus US3. *PLoS Pathog* **15**, e1007559, doi:10.1371/journal.ppat.1007559
624 (2019).
- 625 32 McPherson, J. P. *et al.* Essential role for Bclaf1 in lung development and immune system
626 function. *Cell Death Differ* **16**, 331-339, doi:10.1038/cdd.2008.167 (2009).
- 627 33 Krischuns, T. *et al.* Phosphorylation of TRIM28 Enhances the Expression of IFN-beta
628 and Proinflammatory Cytokines During HPAIV Infection of Human Lung Epithelial
629 Cells. *Front Immunol* **9**, 2229, doi:10.3389/fimmu.2018.02229 (2018).
- 630 34 Hu, C. *et al.* Roles of Kruppel-associated Box (KRAB)-associated Co-repressor KAP1
631 Ser-473 Phosphorylation in DNA Damage Response. *The Journal of biological chemistry*
632 **287**, 18937-18952, doi:10.1074/jbc.M111.313262 (2012).
- 633 35 Wang, Y. *et al.* TRIM28 regulates SARS-CoV-2 cell entry by targeting ACE2. *Cell*
634 *Signal* **85**, 110064, doi:10.1016/j.celsig.2021.110064 (2021).
- 635 36 Tovo, P. A. *et al.* COVID-19 in Children: Expressions of Type I/II/III Interferons,
636 TRIM28, SETDB1, and Endogenous Retroviruses in Mild and Severe Cases. *Int J Mol*
637 *Sci* **22**, doi:10.3390/ijms22147481 (2021).

- 638 37 Paczkowska, M. *et al.* Integrative pathway enrichment analysis of multivariate omics
639 data. *Nature Communications* **11**, 735 (2020).
- 640 38 Knoop, K. *et al.* SARS-coronavirus replication is supported by a reticulovesicular
641 network of modified endoplasmic reticulum. *PLoS Biol* **6**, e226,
642 doi:10.1371/journal.pbio.0060226 (2008).
- 643 39 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**,
644 1260419, doi:10.1126/science.1260419 (2015).
- 645 40 Feng, Z. *et al.* The Novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-
646 CoV-2) Directly Decimates Human Spleens and Lymph Nodes.
647 doi:10.1101/2020.03.27.20045427 (2020).
- 648 41 Yang, A. C. *et al.* Dysregulation of brain and choroid plexus cell types in severe COVID-
649 19. *Nature* **595**, 565-571, doi:10.1038/s41586-021-03710-0 (2021).
- 650 42 Gupta, A. *et al.* Extrapulmonary manifestations of COVID-19. *Nat Med* **26**, 1017-1032,
651 doi:10.1038/s41591-020-0968-3 (2020).
- 652 43 Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res*
653 **47**, D529-D541, doi:10.1093/nar/gky1079 (2019).
- 654 44 Krassowski, M. *et al.* ActiveDriverDB: Interpreting Genetic Variation in Human and
655 Cancer Genomes Using Post-translational Modification Sites and Signaling Networks
656 (2021 Update). *Front Cell Dev Biol* **9**, 626821, doi:10.3389/fcell.2021.626821 (2021).
- 657 45 Popkin, B. M. *et al.* Individuals with obesity and COVID-19: A global perspective on the
658 epidemiology and biological relationships. *Obes Rev* **21**, e13128, doi:10.1111/obr.13128
659 (2020).
- 660 46 Baruah, C., Devi, P. & Sharma, D. K. Sequence Analysis and Structure Prediction of
661 SARS-CoV-2 Accessory Proteins 9b and ORF14: Evolutionary Analysis Indicates Close
662 Relatedness to Bat Coronavirus. *Biomed Res Int* **2020**, 7234961,
663 doi:10.1155/2020/7234961 (2020).
- 664 47 Ma, J. *et al.* The requirement of the DEAD-box protein DDX24 for the packaging of
665 human immunodeficiency virus type 1 RNA. *Virology* **375**, 253-264,
666 doi:10.1016/j.virol.2008.01.025 (2008).
- 667 48 Almasy, K. M., Davies, J. P. & Plate, L. Comparative host interactomes of the SARS-
668 CoV-2 nonstructural protein 3 and human coronavirus homologs. *bioRxiv*,
669 doi:10.1101/2021.03.08.434440 (2021).
- 670 49 Sakai, Y. *et al.* Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral
671 replication. *Virology* **510**, 165-174, doi:10.1016/j.virol.2017.07.019 (2017).
- 672 50 Cottam, E. M. *et al.* Coronavirus nsp6 proteins generate autophagosomes from the
673 endoplasmic reticulum via an omegasome intermediate. *Autophagy* **7**, 1335-1347,
674 doi:10.4161/auto.7.11.16642 (2011).
- 675 51 Sano, H. *et al.* Insulin-stimulated phosphorylation of a Rab GTPase-activating protein
676 regulates GLUT4 translocation. *The Journal of biological chemistry* **278**, 14599-14602,
677 doi:10.1074/jbc.C300063200 (2003).
- 678 52 Stone, S. *et al.* TBC1D1 is a candidate for a severe obesity gene and evidence for a
679 gene/gene interaction in obesity predisposition. *Hum Mol Genet* **15**, 2709-2720,
680 doi:10.1093/hmg/ddl204 (2006).
- 681 53 Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin
682 resistance and type 2 diabetes. *Nature* **512**, 190-193, doi:10.1038/nature13425 (2014).

- 683 54 Coffey, S., Costacou, T., Orchard, T. & Erkan, E. Akt Links Insulin Signaling to
684 Albumin Endocytosis in Proximal Tubule Epithelial Cells. *PLoS One* **10**, e0140417,
685 doi:10.1371/journal.pone.0140417 (2015).
- 686 55 Geraghty, K. M. *et al.* Regulation of multisite phosphorylation and 14-3-3 binding of
687 AS160 in response to IGF-1, EGF, PMA and AICAR. *Biochem J* **407**, 231-241,
688 doi:10.1042/BJ20070649 (2007).
- 689 56 Blasius, M. *et al.* A phospho-proteomic screen identifies substrates of the checkpoint
690 kinase Chk1. *Genome Biol* **12**, R78, doi:10.1186/gb-2011-12-8-r78 (2011).
- 691 57 Zhu, Y. *et al.* A genome-wide CRISPR screen identifies host factors that regulate SARS-
692 CoV-2 entry. *Nat Commun* **12**, 961, doi:10.1038/s41467-021-21213-4 (2021).
- 693 58 Landrum, M. J. *et al.* ClinVar: improvements to accessing data. *Nucleic Acids Res* **48**,
694 D835-D844, doi:10.1093/nar/gkz972 (2020).
- 695 59 Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.
696 *Nucleic Acids Res* **43**, D512-520, doi:10.1093/nar/gku1267 (2015).
- 697 60 UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*
698 **47**, D506-D515, doi:10.1093/nar/gky1049 (2019).
- 699 61 Dinkel, H. *et al.* Phospho.ELM: a database of phosphorylation sites--update 2011.
700 *Nucleic acids research* **39**, D261-267, doi:10.1093/nar/gkq1104 (2011).
- 701 62 Keshava Prasad, T. S. *et al.* Human Protein Reference Database--2009 update. *Nucleic*
702 *acids research* **37**, D767-772, doi:10.1093/nar/gkn892 (2009).
- 703 63 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic
704 variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164,
705 doi:10.1093/nar/gkq603 (2010).
- 706 64 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene
707 Ontology Consortium. *Nature genetics* **25**, 25-29, doi:10.1038/75556 (2000).
- 708 65 Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **46**, D649-
709 D655, doi:10.1093/nar/gkx1132 (2018).
- 710 66 Ruepp, A. *et al.* CORUM: the comprehensive resource of mammalian protein complexes-
711 -2009. *Nucleic acids research* **38**, D497-501, doi:10.1093/nar/gkp914 (2010).
- 712 67 Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler--a web-based toolset
713 for functional profiling of gene lists from large-scale experiments. *Nucleic acids research*
714 **35**, W193-200, doi:10.1093/nar/gkm226 (2007).
- 715 68 Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using
716 g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* **14**, 482-517,
717 doi:10.1038/s41596-018-0103-9 (2019).
- 718 69 Cline, M. S. *et al.* Integration of biological networks and gene expression data using
719 Cytoscape. *Nature protocols* **2**, 2366-2382, doi:10.1038/nprot.2007.324 (2007).

720