

1 **COVID-19: Viral-host interactome analyzed by network based-approach model to study pathogenesis**  
2 **of SARS-CoV-2 infection.**

3

4

5 **Francesco Messina<sup>1°</sup>, Emanuela Giombini<sup>1°</sup>, Chiara Agrati<sup>1</sup>, Francesco Vairo<sup>1</sup>, Tommaso Ascoli**  
6 **Bartoli<sup>1</sup>, Samir Al Moghazi<sup>1</sup>, Mauro Piacentini<sup>1,2</sup>, Franco Locatelli<sup>3</sup>, Gary Kobinger<sup>4</sup>, Markus**  
7 **Maeurer<sup>5</sup>, Alimuddin Zumla<sup>6</sup>, Maria R. Capobianchi<sup>1\*</sup>, Francesco Nicola Lauria<sup>1°</sup>, Giuseppe**  
8 **Ippolito<sup>1°</sup>, COVID 19 INMI Network Medicine for IDs Study Group.**

9 <sup>°</sup> These authors contributed equally to this work.

10

11 <sup>1</sup> National Institute for Infectious Diseases "Lazzaro Spallanzani" IRCCS, Rome, Italy.

12 <sup>2</sup> Department of Biology, University of Rome "Tor Vergata," Rome, Italy.

13 <sup>3</sup> Department of Pediatric Hematology and Oncology, IRCCS Ospedale Pediatrico Bambino Gesù, Piazza  
14 Sant'Onofrio, 4, 00165 Rome, Italy.

15 <sup>4</sup> Département de Microbiologie-Infectiologie et d'Immunologie, Université Laval, Québec City, Québec,  
16 Canada.

17 <sup>5</sup> Champalimaud Centre for the Unknown, Lisbon, Portugal; I. Medizinische Klinik Johannes Gutenberg-  
18 Universität, University of Mainz, 55131 Mainz, Germany.

19 <sup>6</sup> Department of Infection, Division of Infection and Immunity, University College London, and National  
20 Institutes of Health and Research Biomedical Research Centre, University College London Hospitals NHS  
21 Foundation Trust, London, UK.

22

23

24

25 \* **Correspondence** and requests for materials should be addressed to:

26 Maria R. Capobianchi. Electronic address: [maria.capobianchi@inmi.it](mailto:maria.capobianchi@inmi.it)

27

28 **Abstract**

29 **Background.** Epidemiological, virological and pathogenetic characteristics of SARS-CoV-2 infection are  
30 under evaluation. A better understanding of the pathophysiology associated with COVID-19 is crucial to  
31 improve treatment modalities and to develop effective prevention strategies. Transcriptomic and proteomic  
32 data on the host response against SARS-CoV-2 still have anecdotic character; currently available data from  
33 other coronavirus infections are therefore a key source of information.

34 **Methods.** We investigated selected molecular aspects of three human coronavirus (HCoV) infections,  
35 namely SARS-CoV, MERS-CoV and HCoV-229E, through a network based-approach. A functional analysis  
36 of HCoV-host interactome was carried out in order to provide a theoretic host-pathogen interaction model for  
37 HCoV infections and in order to translate the results in prediction for SARS-CoV-2 pathogenesis.

38 The 3D model of S-glycoprotein of SARS-CoV-2 was compared to the structure of the corresponding SARS-  
39 CoV, HCoV-229E and MERS-CoV S-glycoprotein. SARS-CoV, MERS-CoV, HCoV-229E and the host  
40 interactome were inferred through published protein-protein interactions (PPI) as well as gene co-expression,  
41 triggered by HCoV S-glycoprotein in host cells.

42 **Results.** Although the amino acid sequences of the S-glycoprotein were found to be different between the  
43 various HCoV, the structures showed high similarity, but the best 3D structural overlap shared by SARS-  
44 CoV and SARS-CoV-2, consistent with the shared ACE2 predicted receptor. The host interactome, linked to  
45 the S-glycoprotein of SARS-CoV and MERS-CoV, mainly highlighted innate immunity pathway  
46 components, such as Toll Like receptors, cytokines and chemokines.

47 **Conclusions.** In this paper, we developed a network-based model with the aim to define molecular aspects of  
48 pathogenic phenotypes in HCoV infections. The resulting pattern may facilitate the process of structure-  
49 guided pharmaceutical and diagnostic research with the prospect to identify potential new biological targets.

50 **Keywords.** Coronavirus infection; Virus-host interactome; Spike glycoprotein

51

52

53

## 54 **Background**

55 In December 2019, a novel coronavirus (SARS-CoV-2) was first identified as a zoonotic pathogen of  
56 humans in Wuhan, China, causing a respiratory infection with associated bilateral interstitial pneumonia.  
57 The disease caused by SARS-CoV-2 was named by the World Health Organization as COVID-19 and it has  
58 been classified as a global pandemic since it has spread rapidly to all continents. As of March 30, 2020, there  
59 have been 634 835 confirmed COVID-19 cases worldwide with 29 957 deaths reported to the WHO [1].  
60 Whilst clinical and epidemiological data on COVID-19 have become readily available, information on the  
61 pathogenesis of the SARS-CoV-2 infection has not been forthcoming [2].The transcriptomic and proteomic  
62 data on host response against SARS-CoV-2 is scanty and not effective therapeutics and vaccines for COVID-  
63 19 are available yet.

64 Coronaviruses (CoVs) typically affect the respiratory tract of mammals, including humans, and lead to mild  
65 to severe respiratory tract infections [3]. Many emerging HCoV infections have spilled-over from animal  
66 reservoirs, such as HCoV-OC43 and HCoV-229E which cause mild diseases such as common colds [4, 5].  
67 During the past 2 decades, two highly pathogenic HCoVs, severe acute respiratory syndrome coronavirus  
68 (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), have led to global epidemics  
69 with high morbidity and mortality [6]. In this period, a large amount of experimental data associated with the  
70 two infections has allowed to better understand molecular mechanism(s) of coronavirus infection, and  
71 enhance pathways for developing new drugs, diagnostics and vaccines and identification of host factors  
72 stimulating (proviral factors) or restricting (antiviral factors) infection remains poorly understood [7].  
73 Structures of many proteins of SARS-CoV and MERS-CoV, and biological interactions with other viral and  
74 host proteins have been widely explored; through experimental testing of small molecule inhibitors with anti-  
75 viral effects [8, 9]. ACE2, expressed in type 2 alveolar cells in the lung, has been identified as receptor of  
76 SARS-CoV and SARS-CoV-2, while dipeptidyl peptidase DPP4 was identified as the specific receptor for  
77 MERS-CoV [10, 11].

78 The investigation of structural genomics and interactomics of SARS-CoV-2 can be implemented through an  
79 integrated bioinformatics approach. Structural analysis of specific SARS-CoV-2 proteins, in particular Spike  
80 glycoproteins (S-glycoproteins), and their interactions with human proteins, can guide the identification of  
81 the putative functional sites and help to better define the pathologic phenotype of the infection. This

82 functional interaction analysis between the host and other HCoV, combined with an evolutionary sequence  
83 analysis of SARS-CoV-2, can be used to guide new treatment and prevention interventions.

84 We investigated here biologically and clinically relevant molecular targets of three human coronaviruses  
85 (HCoV) infections using a network based approach. A functional analysis of HCoV-host interactome was  
86 carried out in order to provide a theoretic host-pathogen interaction model for HCoV infections, and to  
87 predict viable models for SARS-CoV-2 pathogenesis. Three HCoV causing respiratory diseases were used as  
88 the model targets, namely: SARS-CoV, that shares with SARS-CoV-2 a strong genetic similarity, including  
89 MERS-CoV, and HCoV-229E.

90

## 91 **Methods**

### 92 *Comparative reconstruction of S-glycoprotein in HCVs*

93 The reconstruction of virus-host interactome was carried out using the RWR algorithm to explore the human  
94 PPI network and the multilayer PPI platform enriched with gene expression data sets. 259 sequences of  
95 CoVs, infecting different animal hosts (Table S1 in Additional file 1), were downloaded by GSAID and  
96 NCBI database in order to evaluate the variability in the S gene. SARS-CoV, HCoV-229E and MERS-CoV  
97 and other CoV full genome sequence groups were aligned with MAFFT [12], synonymous and non-  
98 synonymous mutations, and amino acid similarity were calculated using the SSE program with a sliding  
99 windows of 250 nucleotides and a pass of 25 nu [13]. A homology model was built for the amino acid  
100 sequences of the S-glycoprotein, derived from the full genome sequence obtained at “SARS-CoV-  
101 2/INMI1/human/2020/ITA” (MT066156.1). The Swiss pdb server was used to construct three-dimensional  
102 models for the S-glycoprotein of SARS-CoV-2 [14]. Among proteins with a 3D structure, the best match  
103 with the “SARS-CoV-2/INMI1/human/2020/ITA” sequence was the 6VSB.1, that was evaluated considering  
104 the identity of two amino acid sequences and the QMEN value included in Swiss pdb server. The model of a  
105 single chain was overlapped with the three-dimensional structure of S-glycoprotein single chain belonging to  
106 SARS-CoV (5WRG), HCoV-229E (6U7H.1) and MERS-CoV (5X59), using Chimera 1.14 [15]. In order to  
107 better evaluate the conservation of the sequence in each site, all sequences were aligned with MAFFT and

108 the topology of all structures were compared. The detailed description of the reconstruction of S-  
109 glycoprotein structure is reported in the Appendix.

110

### 111 ***PPI and Gene Co-expression Network***

112 Network analysis, based on protein-protein interactions and gene expression data, was performed in order to  
113 view all possible virus-host protein interactions during the HCoV infections. Since the SARS-CoV-2 genome  
114 exhibits substantial similarity to the SARS-CoV genome [16] and subsequently also the proteome [17], we  
115 hypothesized that several molecular interactions that were observed in the SARS-CoV interactome will be  
116 preserved in the SARS-CoV-2 interactome. Virus-Host interactomes (SARS-CoV, MERS-CoV, HCoV-  
117 229E) were inferred through published PPI data, using two publicly accessible databases (STRING Viruses  
118 and VirHostNet), as well as published scientific reports with a focus on virus-host interactions [18-20]. As a  
119 next step, the virus-host PPI list, extracted in this first step, was merged with additional PPI databases, i.e.  
120 BioGrid, InnateDB-All, IMEx, IntAct, MatrixDB, MBInfo, MINT, Reactome, Reactome-FIs, UniProt,  
121 VirHostNet, BioData, CCSB Interactome Database, using R packages PSICQUIC and biomaRt [21, 22]. In  
122 total, a large PPI interaction database was assembled, including 13,020 nodes and 71,496 interactions.

123 The gene expression data set was built from the Protein Atlas database, using tissue and cell line data [23].  
124 To identify the most likely interactions, and to obtain functional information, Random walk with restart  
125 (RWR), a state-of-the-art guilt-by-association approach by R package RandomWalkRestartMH [24] was  
126 used. It allows to establish a proximity network from a given protein (seed), to study its functions, based on  
127 the premise that nodes related to similar functions tend to lie close to each other in the network. For each  
128 node, a score was computed as measure of proximity to the seed protein. S-glycoproteins of SARS-CoV,  
129 MERS-CoV and HCoV-229E were used as seed in the application of the RWR algorithm.

130

### 131 ***Functional Enrichment Analysis***

132 To evaluate functional pathways of proteins involved in host response, gene enrichment analysis was  
133 performed, using Kyoto Encyclopedia of Genes and Genomes (KEGG) human pathways and Gene Ontology  
134 databases. Network representation from the gene enrichment analysis was performed using ShinyGO  
135 v0.61[25]. The statistical significance was obtained, calculating the False Discovery Rate (FDR).

## 136 **Results**

### 137 *Structure of S-glycoprotein CoVs*

138 To evaluate the diversity along the full genome, pairwise distance was calculated on 259 HCoV genomes.  
139 Diversity was distributed along the entire CoV genome, with the most conserved region located in Orf1ab, as  
140 expected, while the spike gene region exhibited a rather high diversity (Figure S1 in Appendix).  
141 Consequently, the analysis was focused on the S-glycoprotein, as a key virus component involved in host  
142 interaction [26]. A 3D model of S-glycoprotein of the SARS-CoV-2 sequence (MT066156.1) was built on  
143 the sequence obtained at Laboratory of Virology, National Institute of Infectious Diseases "L. Spallanzani"  
144 IRCCS, using Swiss pdb viewer server (Figure S2 a/b in Appendix). The SARS-CoV-2 S-glycoprotein  
145 structure was then compared to other HCoVs as shown in Figure S2 in Appendix. The S-glycoprotein  
146 structures of the various HCoV were very similar overall. In particular, a strong similarity was shown in the  
147 RBD (nCoV: residues 319–591) [27], and this was most evident for the comparison between SARS-CoV-2  
148 and SARS-CoV, which share the same cell receptor (ACE2). The amino acid differences among the S-  
149 glycoproteins of the selected HCoVs (SARS-CoV-2, MERS-CoV, SARS-CoV, HCoV-229E) are shown in  
150 Figure S3 in Appendix, where a lower topology similarity was observed with HCoV-229E S-glycoprotein,  
151 which binds a different host receptor.  
152 Overall, the pattern arising from such comparison was consistent with specific host receptors, as well as with  
153 different host reservoirs and ancestry [28].

154

### 155 *Human CoV and host interactome*

156 An interactome map was built to highlight biological connections among S-glycoprotein and the human  
157 proteome. Using the analysis pipeline described in the methods, a large PPI interaction database was  
158 assembled, including 13,020 nodes and 71,496 interactions between human host and the three selected  
159 viruses (SARS-CoV, MERS-CoV and HCoV-229E).

160 The interactome reconstruction was obtained with the RWR analysis, finding 200 closest proteins to seed, or  
161 S-glycoproteins of HCoV-229E, SARS-CoV and MERS-CoV (Figures S4-S6 in Appendix). In Table S2, S3  
162 and S4 in Additional file 1, lists of genes selected by RWR algorithm for HCoV-229E, SARS-CoV and  
163 MERS-CoV, along with proximity score were reported. In order to further dissect the S-glycoprotein-host

164 interactions, enrichment analysis was carried out with Reactome and KEGG databases. Reactome pathway  
165 enrichment analysis revealed biological pathways of DNA repair, transcription and gene regulation for the  
166 HCoV-229E S-glycoprotein, with high significance (FDR <0.01%). KEGG pathway enrichment analysis  
167 revealed ubiquitin mediated proteolysis as the most significant pathway (FDR <0.01%), as well as altered  
168 cellular proliferation pathways, associated with other viral infections (Hepatitis B, measles, Epstein-Barr  
169 virus infection and Human T-cell leukemia virus 1 infection) as well as with carcinogenesis (Figure 1). Next,  
170 the RWR algorithm was applied to a multilayer network built on the PPI interactome and on the Gene  
171 Coexpression (COEX) network, again with S-glycoprotein of HCoV-229E as seed. The results highlighted a  
172 set of genes that are connected in both PPI and COEX analysis, including ANPEP, RAD18, APEX, POLH,  
173 APEX1, TERF2, RAD51, CDC7, USP7, XRCC5, RAD18, FEN1, PCNA, all associated to the GO biological  
174 process category of DNA repair (FDR< 0.0001%) (Figure 2). The same analyses were conducted for SARS-  
175 CoV and MERS-CoV.

176

177 The Reactome pathway enrichment analysis for the SARS-CoV revealed S-glycoprotein connection with  
178 early activation of innate immune system, such as the Toll Like Receptor Cascade and TGF- $\beta$ , with a strong  
179 significance (FDR < 0.0001%), while the KEGG pathway enrichment analysis revealed an association with  
180 cellular proliferation, TGF- $\beta$  and other infection-related pathways (FDR<0.0001%) (Figure 3). The PPI-  
181 COEX multilayer analysis highlighted a set of genes that are connected in both PPI and COEX analysis,  
182 i.e.CLEC4G, CLEC4M, CD209, ACE2, RPSA, all associated to the GO biological process category of  
183 SARS-CoV entry into host cell (FDR < 0.01%) (Figure 4).

184

185

186

187 In MERS-CoV, the Reactome pathway enrichment analysis showed a strong association with membrane  
188 signals activated by GPCR ligand binding (FDR < 0.0001%), and chemokine/chemokine receptor pathways.  
189 Consistent results were obtained with KEGG pathway enrichment, that highlighted cytokine-cytokine  
190 receptor and chemokine signalling pathways (FDR<0.0001%) (Figure 5). Finally, PPI-COEX multilayer  
191 analysis evidenced, for both PPI and COEX, CCR4, CXCL2, CXCL10, CXCL9, PF4, PF4V1, CCL11,

192 CXCL11, XCL1, CXCR4 and CXCL14, all genes identified by the GO biological processes involved in the  
193 chemokine cascade (FDR < 0.0001%), in line with the results obtained with enrichment analyses (Figure 6).

194

## 195 **Discussion**

### 196 *In-depth comparative analysis of S-glycoprotein*

197 We applied network analysis, based on protein-protein interactions and gene expression data, in order to  
198 describe the interactome of the coronavirus S-glycoprotein and host proteins, with the aim to better  
199 understand SARS-CoV-2 pathogenesis. A preliminary structural analysis was conducted on the S-  
200 glycoprotein of SARS-CoV-2 as compared to the other 3 HCoV, using the S-glycoprotein as a model to shed  
201 light on the host-pathogen interaction in the dynamic process of SARS-CoV-2 infection. Although the amino  
202 acid sequences of the S-glycoprotein were different between the various HCoVs, the structural analysis  
203 exhibited high similarity; the best 3D structural overlap was found for SARS-CoV and SARS-CoV-2,  
204 consistent with the shared ACE2 predicted receptor.

205 Of note, the newly discovered SARS-CoV-2 genome has revealed differences between SARS-CoV-2 and  
206 SARS or SARS-like coronaviruses [28]. Although no amino acid substitutions were present in the receptor-  
207 binding motifs, that directly interact with human receptor ACE2 protein in SARS-CoV, six mutations  
208 occurred in the other region of the RBD [28, 29] were identified. On the other hand, the genomic  
209 comparative analysis highlighted the a strong diversity in the S gene among CoV in different hosts,  
210 confirming the biologically vital role of the S-glycoprotein as a key factor in viral entry in cross-species  
211 transmission events [30].

212 In addition, the comparative 3D structural data may facilitate the definition of already known antibody  
213 epitopes in the S-glycoprotein of other coronaviruses, it will also be useful in rational vaccine design and in  
214 gauging anti-virus directed immune responses after vaccination [27]. In fact, S-glycoprotein remains an  
215 important target for vaccines and drugs. previously evaluated in SARS and MERS, while a neutralizing  
216 antibody targeting the S-glycoprotein protein could provide passive immunity. The host interactome, linked  
217 to S-glycoprotein of SARS-CoV and MERS-CoV, mainly highlighted innate immunity pathway components,  
218 such as Toll Like receptors, cytokines and chemokines. The 3D structural analysis confirmed that we



219 established that S-glycoprotein of SARS-CoV-2 has strong similarity in the 3D structure with SARS-CoV  
220 [16].

221

222

### 223 *Host interactome in all three HCoV infections*

224 The reconstruction of virus-host interactome was carried out using RWR algorithm to explore the human PPI  
225 network and studying PPI and COEX multilayer. The PPI network topology of host interactome in all three  
226 infections indicated the presence of several hub proteins. In the HCoV-229E - host interactome hub position  
227 was hold by RAD18 and APEX, which play an important role in DNA repair due to UV damage in phase S  
228 [31].

229 For the SARS-CoV interactome, the gene hubs were identified in ACE2, CLEC4G and CD209, which are  
230 known interactors with S-glycoprotein of SARS-CoV [32, 33].

231 In fact, two independent mechanisms were described as trigger of SARS-CoV infection: proteolytic cleavage  
232 of ACE2 and cleavage of S-glycoprotein. The latter activates the glycoprotein for cathepsin L-independent  
233 host cell entry. Activated the S-glycoprotein by cathepsin L mechanism in host cell entry was reported in  
234 many infection of CoV, such as HCoV-229E and SARS-CoV [34, 35]. A recent study speculated that this  
235 interaction will be preserved in SARS-CoV-2 [17], but might be disrupted of a substantial number of  
236 mutations in the receptor binding site of S gene will occur. Likewise, the S-glycoprotein in SARS-CoV-2 is  
237 expected to interact with type II transmembrane protease (TMPRSS2) and probably is involved in inhibition  
238 of antibody-mediated neutralization [36, 37]. It is rather unexpected that, for this virus, no intracellular  
239 pathways were highlighted by the multilayer analysis, suggesting that this field is still open to further  
240 investigation.

241 In MERS-CoV infection a gene hub role was described for DPP4, which is known to regulate cytokine levels  
242 through catalytic cleavage [38]. Immune cell-recruiting chemokines and cytokines, such as IP-10/CXCL-10,  
243 MCP-1/CCL-2, MIP-1 $\alpha$ /CCL-3, RANTES/CCL-5, can be strongly induced by MERS-CoV, showing higher  
244 inducibility in human monocyte-derived macrophages by MERS-CoV as compared to than SARS-CoV  
245 infection [39].

246 Finally, biological pathways, revealed by enrichment analysis in over all models, supported early activation  
247 of innate immune system, as Toll Like receptor Cascade and TGF- $\beta$  for SARS-CoV, or chemokine and  
248 cytokine pathways and infection-related pathways for MERS-CoV, with a strong significance for both.

249

250

### 251 *Pathogenic model for HCoV infections*

252 We constructed a host molecular interactome with SARS-CoV, MERS-CoV and HCoV-229E in patients  
253 with cancer, assuming that most of these interactions, especially for SARS-CoV, are shared with SARS-  
254 CoV-2. A network-based methodology, along with guilt-by-association algorithm (RWR), was applied to  
255 define the pathological model of COVID-19 and provide a treatment of SARS-CoV-2, using existing  
256 transcriptomic and proteomic information.

257 Based on the main pathways identified by the network-based interactome analysis, the following issues  
258 require focus further study:

259 **First**, The predicted receptor for SARS-CoV-2 has been inferred to be ACE-2, i.e. the same used by SARS-  
260 CoV, based on the high similarity of the S-glycoprotein of the two viruses, and this is the basis for  
261 hypothesizing to use SARS-CoV as a model for virus-host interactome in COVID-19;

262 **Second**, Mitogen activated protein kinase (MAPK) is a major cell signalling pathway that is known to be  
263 activated by diverse groups of viruses, and plays an important role in cellular response to viral infections.  
264 MAPK interacting kinase 1 (MNK1) has been shown to regulate both cap-dependent and internal ribosomal  
265 entry sites (IRES)-mediated mRNA translation;

266 **Third**, The identification of the MAPK pathway in SARS-CoV model is highly consistent with in vivo  
267 model, where P38 MAPK was found increased in the lungs of mice infected with SARS-CoV [40];

268 **Fourth**, The identification of the TGF- $\beta$  pathway in S-glycoprotein-induced interactome for SARS-CoV of  
269 particular interest, due to the previous evidence that this virus, and in particular its protease, triggered the  
270 TGF- $\beta$  through the p38 MAPK/STAT3 pathway in alveolar basal epithelial cells [41, 42];

271 **Fifth**, Innate immune pathways were identified in S-glycoprotein-induced models of SARS-CoV and  
272 MERS-CoV, as Toll Like receptor, cytokine and chemokine.

273 Every described pathway can be matched with clinical aspects, the data presented in this report may  
274 therefore aid to design a ‘blue print’ for SARS-CoV-2 associated pathogenicity.

275 The severity and the clinical picture of SARS-CoV and MERS-CoV infections could be related to the  
276 activation of exaggerated immune mechanism, causing uncontrolled inflammation [43]; however, the role of  
277 strong immune response in SARS-CoV-2 infection severity is still uncertain.

278 However, we may consider that host kinases link multiple signalling pathways in response to a broad array  
279 of stimuli, including viral infections. TGF- $\beta$ , produced during the inflammatory phase by macrophages, is an  
280 important mediator of fibroblast activation and tissue repair. High levels of systemic inflammatory  
281 cytokines/chemokines has been widely reported for MERS-CoV infections [44-46], correlating with  
282 immunopathology and massive pulmonary infiltration into the lungs [47]. Also the HCoV-229E infection  
283 can be described with this distance model, although this infection was not associated with a severe  
284 respiratory disease. In fact, HCoV-229E is responsible for mild upper respiratory tract infections, such as  
285 common colds, with only occasional spreading to the lower respiratory tract, but it interacts with dendritic  
286 cells in the upper respiratory tract, inducing a cytopathic effect [48].

287

288

289

## 290 **Conclusions**

291 In conclusion, we developed a network-based model, which could be the framework for structure-guided  
292 research process and for the pathogenetic evaluation of specific clinical outcome. Accurate structural 3D  
293 protein models and their interaction with host receptor proteins can allow to build a more detailed theoretical  
294 disease model for each HCoV infection, and support the drawing of a disease model for COVID-19. Our  
295 analyses suggests it is important to carry out *in silico* experiments and simulations through specific  
296 algorithms.

297

## 298 **Limitations to our study**

299 A single protein, namely S-glycoprotein was used as seed, therefore the highlighted interactions were limited  
300 to those connected with this unique viral protein. However, this is a proof of concept study, from which it  
301 appears that a similar approach may be used to study other viral proteins interacting with host cell pathways.  
302 Another limitation is that the pathway analysis did not consider tissue and cell type diversity. Finally, the low  
303 threshold established for the number of nodes found by RWR (200) limited the reconstruction of the entire  
304 pathways. However, this was a software-imposed threshold.  
305 In summary, the interactome analysis aided to guide the design of novel models of SARS-CoV  
306 pathogenicity.

307

### 308 **Abbreviations**

309 CoVs: Coronaviruses; HCoV: Human Coronavirus; PPI: Protein-Protein Interactions; COEX: Gene  
310 Coexpression data; SARS-CoV: Severe Acute Respiratory Syndrome Coronavirus; MERS-CoV: Middle  
311 East Respiratory Syndrome Coronavirus; S-glycoprotein: Spike glycoprotein; RWR: Random walk with  
312 restart; FDR: False Discovery Rate.

313

### 314 **Authors' contributions**

315 Conceptualization: F.M., E.G., F.N.L. Data curation: F.M., E.G., Formal analysis: F.M., E.G., Funding  
316 acquisition: M.R.C., G.I., F.V. Investigation: F.M., E.G., F.N.L. Methodology: F.M., E.G., T.A., S.A.  
317 Resources: M.R.C., G.I. Software: F.M., E.G. Supervision: M.R.C., F.N.L. Validation: C.A., F.V., G.K.,  
318 M.M., M.P. Visualization: F.M., E.G. Writing ± original draft: F.M., E.G., M.R.C., F.N.L., G.K., A.Z., M.M.  
319 Writing ± review & editing: F.M., E.G., M.R.C., F.N.L. A.Z. All authors reviewed and approved the  
320 manuscript.

321

322 1

323

324

325

326 **Acknowledgments**

327 We gratefully acknowledge: Collaborators Members of INMI COVID-19 study group; COVID 19 INMI  
328 Network Medicine for IDs Study Group:  
329 Abbate Isabella, Agrati Chiara, Al Moghazi Samir, Ascoli Bartoli Tommaso, Capobianchi Maria Rosaria,  
330 Capone Alessandro, Goletti Delia, Rozera Gabriella, Nisii Carla, Gagliardini Roberta, Ciccocanti Fabiola,  
331 Fimia Gian Maria, Nicastrì Emanuele, Giombini Emanuela, Lanini Simone, D'Abramo Alessandra,  
332 Rinonapoli Gabriele, Girardi Enrico, Bartolini Barbara, Montaldo Chiara, Marconi Raffaella, Antonio Addis,  
333 Bradley Maron, Bianconi Ginestra, De Meulder Bertrand, Kennedy Jason, Khader Shabaana Abdul, Luca  
334 Francesca, Maeurer Markus, Piacentini Mauro, Merler Stefano, Pantaleo Giuseppe, Rafick-Pierre Sekaly,  
335 Sanna Serena, Segata Nicola, Zumla Alimuddin, Messina Francesco, Lauria Francesco Nicola, Ippolito  
336 Giuseppe, Vairo Francesco.

337

338 **Competing interests**

339 The authors declare no competing interests.

340

341 **Consent for publication**

342 Not applicable.

343

344 **Ethics approval and consent to participate**

345 Not applicable.

346

347 **Funding**

348 INMI authors are supported by the Italian Ministry of Health (Ricerca Corrente Linea 1). G. Ippolito and A.  
349 Zumla are co-principal investigator of the Pan-African Network on Emerging and Re-emerging Infections  
350 (PANDORA-ID-NET), funded by the European & Developing Countries Clinical Trials Partnership,  
351 supported under Horizon 2020. Sir Zumla is in receipt of a National Institutes of Health Research senior

352 investigator award. M. Maeurer is a member of the innate immunity advisory group of the Bill & Melinda  
353 Gates Foundation, and is funded by the Champalimaud Foundation.

354

### 355 **Availability of data and materials**

356 PPI data of SARS-CoV, MERS-CoV, HCoV-229E S-glycoprotein were inferred through published PPI data,  
357 using STRING Viruses (<http://viruses.string-db.org/>) and VirHostNet (<http://virhostnet.prabi.fr/>), as well as  
358 published scientific reports with a focus on virus-host interactions [18-20]. Human PPI databases (BioGrid,  
359 InnateDB-All, IMEx, IntAct, MatrixDB, MBInfo, MINT, Reactome, Reactome-FIs, UniProt, VirHostNet,  
360 BioData, CCSB Interactome Database), using R packages PSICQUIC  
361 (<https://bioconductor.org/packages/release/bioc/html/PSICQUIC.html>) and biomaRt  
362 (<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>) [21, 22]. The gene expression data set  
363 was built from the Protein Atlas database (<https://www.proteinatlas.org/>) [23].

364

365

### 366 **References**

367

- 368 1. Johns Hopkins University. global-cases-covid-19. 2020; Available from:  
369 <https://www.gisaid.org/epiflu-applications/global-cases-covid-19/>.
- 370 2. Jernigan DB. Update: Public Health Response to the Coronavirus Disease 2019 Outbreak - United  
371 States, February 24, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(8):216-9. doi:  
372 10.15585/mmwr.mm6908e1.
- 373 3. Paules CI, Marston HD, Fauci AS. Coronavirus Infections-More Than Just the Common Cold.  
374 JAMA. 2020. doi: 10.1001/jama.2020.0757.
- 375 4. Walsh EE, Shin JH, Falsey AR. Clinical impact of human coronaviruses 229E and OC43 infection in  
376 diverse adult populations. J Infect Dis. 2013;208(10):1634-42. doi: 10.1093/infdis/jit393.
- 377 5. Hendley JO, Fishburne HB, Gwaltney JM, Jr. Coronavirus infections in working adults. Eight-year  
378 study with 229 E and OC 43. Am Rev Respir Dis. 1972;105(5):805-11. doi: 10.1164/arrd.1972.105.5.805.

- 379 6. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into  
380 emerging coronaviruses. *Nat Rev Microbiol.* 2016;14(8):523-34. doi: 10.1038/nrmicro.2016.81.
- 381 7. de Wilde AH, Wannee KF, Scholte FE, Goeman JJ, Ten Dijke P, Snijder EJ, et al. A Kinome-Wide  
382 Small Interfering RNA Screen Identifies Proviral and Antiviral Host Factors in Severe Acute Respiratory  
383 Syndrome Coronavirus Replication, Including Double-Stranded RNA-Activated Protein Kinase and Early  
384 Secretory Pathway Proteins. *J Virol.* 2015;89(16):8318-33. doi: 10.1128/JVI.01029-15.
- 385 8. Cotten M, Watson SJ, Kellam P, Al-Rabeeh AA, Makhdoom HQ, Assiri A, et al. Transmission and  
386 evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study.  
387 *PLoS One.* 2013;382(9909):1993-2002. doi: 10.1016/S0140-6736(13)61887-5.
- 388 9. Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, et al. Cryo-EM structures of MERS-CoV and SARS-  
389 CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun.* 2017;8:15092. doi:  
390 10.1038/ncomms15092.
- 391 10. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated  
392 with a new coronavirus of probable bat origin. *Nature.* 2020. doi: 10.1038/s41586-020-2012-7.
- 393 11. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with  
394 Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-33. doi: 10.1056/NEJMoa2001017.
- 395 12. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple  
396 sequence alignments. *BMC Bioinformatics.* 2018;34(14):2490-2. doi: 10.1093/bioinformatics/bty121.
- 397 13. Simmonds P. SSE: a nucleotide and amino acid sequence analysis platform. *BMC Res Notes.*  
398 2012;5:50. doi: 10.1186/1756-0500-5-50.
- 399 14. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL:  
400 homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296-W303.  
401 doi: 10.1093/nar/gky427.
- 402 15. Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE. Tools for integrated sequence-structure  
403 analysis with UCSF Chimera. *BMC Bioinformatics.* 2006;7:339. doi: 10.1186/1471-2105-7-339.
- 404 16. Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. *J Virol.*  
405 2007;81(1):20-9. doi: 10.1128/JVI.01358-06.

- 406 17. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from  
407 Wuhan: An analysis based on decade-long structural studies of SARS. *J Virol.* 2020. doi:  
408 10.1128/JVI.00127-20.
- 409 18. Cook HV, Doncheva NT, Szklarczyk D, von Mering C, Jensen LJ. Viruses.STRING: A Virus-Host  
410 Protein-Protein Interaction Database. *Viruses.* 2018;10(10). doi: 10.3390/v10100519.
- 411 19. Letko M, Miazgowicz K, McMinn R, Seifert SN, Sola I, Enjuanes L, et al. Adaptive Evolution of  
412 MERS-CoV to Species Variation in DPP4. *Cell Rep.* 2018;24(7):1730-7. doi: 10.1016/j.celrep.2018.07.045.
- 413 20. Pfefferle S, Schopf J, Kogl M, Friedel CC, Muller MA, Carbajo-Lozoya J, et al. The SARS-  
414 coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS*  
415 *Pathog.* 2011;7(10):e1002331. doi: 10.1371/journal.ppat.1002331.
- 416 21. Aranda B, Blankenburg H, Kerrien S, Brinkman FS, Ceol A, Chautard E, et al. PSICQUIC and  
417 PSISCOPE: accessing and scoring molecular interactions. *Nat Methods.* 2011;8(7):528-9. doi:  
418 10.1038/nmeth.1637.
- 419 22. Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, et al. BioMart--biological  
420 queries made easy. *BMC Genomics.* 2009;10:22. doi: 10.1186/1471-2164-10-22.
- 421 23. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhorji G, et al. A pathology atlas of the  
422 human cancer transcriptome. 2017;357(6352). doi: 10.1126/science.aan2507.
- 423 24. Valdeolivas A, Tichit L, Navarro C, Perrin S, Odelin G, Levy N, et al. Random walk with restart on  
424 multiplex and heterogeneous biological networks. 2019;35(3):497-505. doi: 10.1093/bioinformatics/bty637.
- 425 25. Ge SX, Jung D, Yao R. ShinyGO: a graphical enrichment tool for animals and plants. 2019. doi:  
426 10.1093/bioinformatics/btz931.
- 427 26. Gallagher TM, Buchmeier MJ. Coronavirus spike proteins in viral entry and pathogenesis. *Virology.*  
428 2001;279(2):371-4. doi: 10.1006/viro.2000.0757.
- 429 27. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the  
430 2019-nCoV spike in the prefusion conformation. 2020;367(6483):1260-3. doi: 10.1126/science.abb2507.
- 431 28. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome Composition and Divergence of  
432 the Novel Coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe.* 2020. doi:  
433 10.1016/j.chom.2020.02.001.



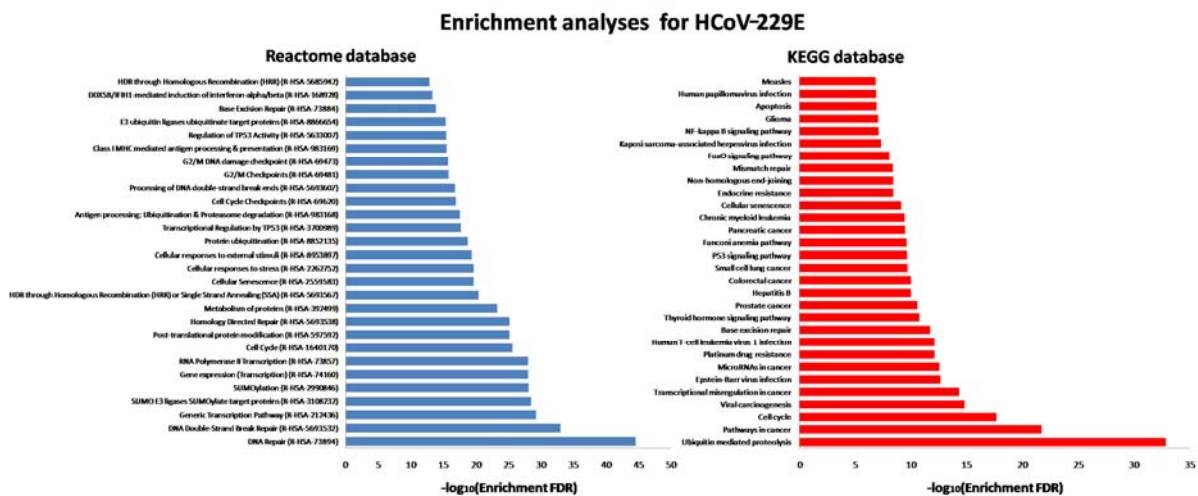
- 434 29. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a  
435 bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*. 2013;503(7477):535-8. doi:  
436 10.1038/nature12711.
- 437 30. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of  
438 coronavirus cross-species transmission. *J Virol*. 2010;84(7):3134-46. doi: 10.1128/JVI.01394-09.
- 439 31. Srivastava M, Chen Z, Zhang H, Tang M, Wang C, Jung SY, et al. Replisome Dynamics and Their  
440 Functional Relevance upon DNA Damage through the PCNA Interactome. *Cell Rep*. 2018;25(13):3869-83  
441 e4. doi: 10.1016/j.celrep.2018.11.099.
- 442 32. Gramberg T, Hofmann H, Moller P, Lalor PF, Marzi A, Geier M, et al. LSECtin interacts with  
443 filovirus glycoproteins and the spike protein of SARS coronavirus. *Virology*. 2005;340(2):224-36. doi:  
444 10.1016/j.virol.2005.06.026.
- 445 33. Yang ZY, Huang Y, Ganesh L, Leung K, Kong WP, Schwartz O, et al. pH-dependent entry of severe  
446 acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell  
447 transfer through DC-SIGN. *J Virol*. 2004;78(11):5642-50. doi: 10.1128/JVI.78.11.5642-5650.2004.
- 448 34. Bertram S, Dijkman R, Habjan M, Heurich A, Gierer S, Glowacka I, et al. TMPRSS2 activates the  
449 human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the  
450 respiratory epithelium. *J Virol*. 2013;87(11):6150-60. doi: 10.1128/JVI.03372-12.
- 451 35. Bosch BJ, Bartelink W, Rottier PJ. Cathepsin L functionally cleaves the severe acute respiratory  
452 syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. *J Virol*.  
453 2008;82(17):8887-90. doi: 10.1128/JVI.00415-08.
- 454 36. Glowacka I, Bertram S, Muller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2  
455 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces  
456 viral control by the humoral immune response. *J Virol*. 2011;85(9):4122-34. doi: 10.1128/JVI.02232-10.
- 457 37. Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, Erichsen S, et al. SARS-CoV-2  
458 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*.  
459 2020. doi: 10.1016/j.cell.2020.02.052.
- 460 38. Zhong J, Rajagopalan S. Dipeptidyl Peptidase-4 Regulation of SDF-1/CXCR4 Axis: Implications for  
461 Cardiovascular Disease. *Front Immunol*. 2015;6:477. doi: 10.3389/fimmu.2015.00477.

- 462 39. Zhou J, Chu H, Li C, Wong BH, Cheng ZS, Poon VK, et al. Active replication of Middle East  
463 respiratory syndrome coronavirus and aberrant induction of inflammatory cytokines and chemokines in  
464 human macrophages: implications for pathogenesis. *J Infect Dis.* 2014;209(9):1331-42. doi:  
465 10.1093/infdis/jit504.
- 466 40. Jimenez-Guardeno JM, Nieto-Torres JL, DeDiego ML, Regla-Nava JA, Fernandez-Delgado R,  
467 Castano-Rodriguez C, et al. The PDZ-binding motif of severe acute respiratory syndrome coronavirus  
468 envelope protein is a determinant of viral pathogenesis. *PLoS Pathog.* 2014;10(8):e1004320. doi:  
469 10.1371/journal.ppat.1004320.
- 470 41. Li SW, Wang CY, Jou YJ, Yang TC, Huang SH, Wan L, et al. SARS coronavirus papain-like  
471 protease induces Egr-1-dependent up-regulation of TGF-beta1 via ROS/p38 MAPK/STAT3 pathway. *Sci*  
472 *Rep.* 2016;6:25754. doi: 10.1038/srep25754.
- 473 42. Li SW, Yang TC, Wan L, Lin YJ, Tsai FJ, Lai CC, et al. Correlation between TGF-beta1 expression  
474 and proteomic profiling induced by severe acute respiratory syndrome coronavirus papain-like protease.  
475 *Proteomics.* 2012;12(21):3193-205. doi: 10.1002/pmic.201200225.
- 476 43. Newton AH, Cardani A, Braciale TJ. The host immune response in respiratory virus infection:  
477 balancing virus clearance and immunopathology. *Semin Immunopathol.* 2016;38(4):471-82. doi:  
478 10.1007/s00281-016-0558-0.
- 479 44. Kindler E, Thiel V, Weber F. Interaction of SARS and MERS Coronaviruses with the Antiviral  
480 Interferon Response. *Adv Virus Res.* 2016;96:219-43. doi: 10.1016/bs.aivir.2016.08.006.
- 481 45. Mahallawi WH, Khabour OF, Zhang Q, Makhdom HM, Suliman BA. MERS-CoV infection in  
482 humans is associated with a pro-inflammatory Th1 and Th17 cytokine profile. *Cytokine.* 2018;104:8-13. doi:  
483 10.1016/j.cyto.2018.01.025.
- 484 46. Tynell J, Westenius V, Ronkko E, Munster VJ, Melen K, Osterlund P, et al. Middle East respiratory  
485 syndrome coronavirus shows poor replication but significant induction of antiviral responses in human  
486 monocyte-derived macrophages and dendritic cells. *J Gen Virol.* 2016;97(2):344-55. doi:  
487 10.1099/jgv.0.000351.

488 47. Mella C, Suarez-Arrabal MC, Lopez S, Stephens J, Fernandez S, Hall MW, et al. Innate immune  
 489 dysfunction is associated with enhanced disease severity in infants with severe respiratory syncytial virus  
 490 bronchiolitis. *J Infect Dis*. 2013;207(4):564-73. doi: 10.1093/infdis/jis721.

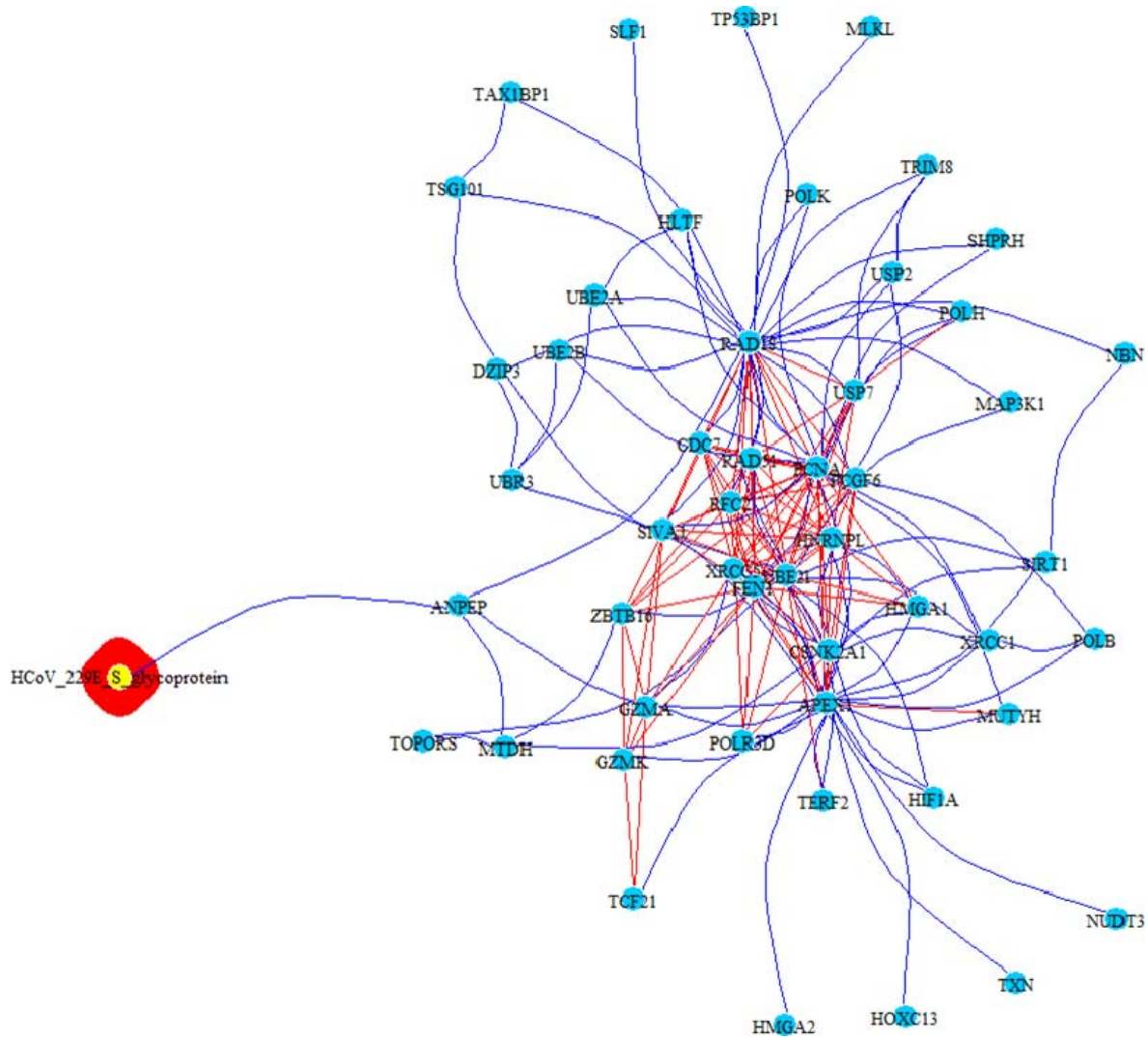
491 48. Mesel-Lemoine M, Millet J, Vidalain PO, Law H, Vabret A, Lorin V, et al. A human coronavirus  
 492 responsible for the common cold massively kills dendritic cells but not monocytes. *J Virol*.  
 493 2012;86(14):7577-87. doi: 10.1128/JVI.00269-12.

494  
 495  
 496  
 497  
 498  
 499



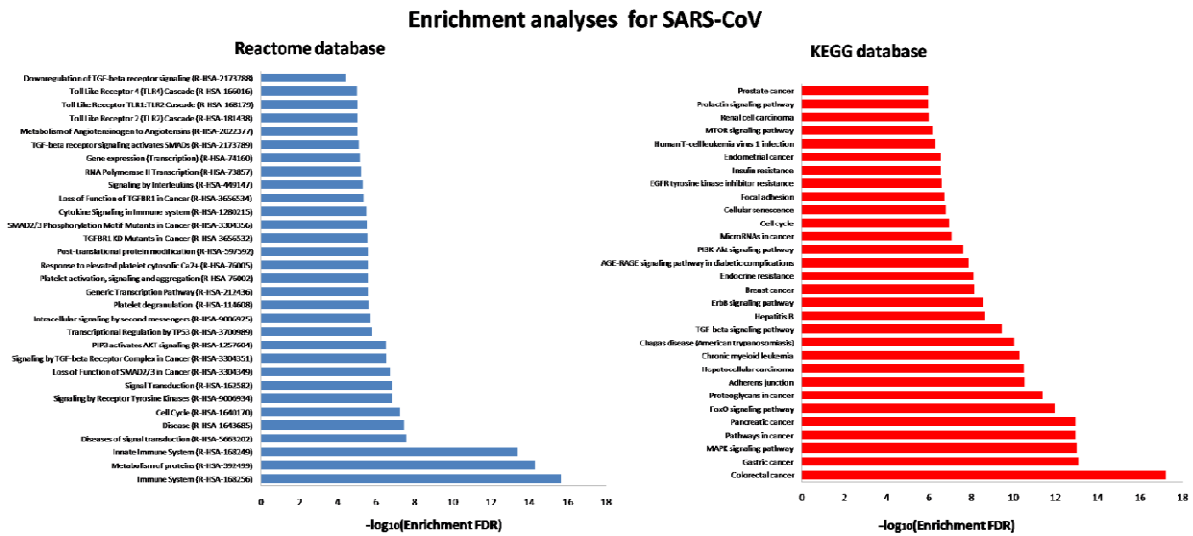
500  
 501  
 502

Figure 1. KEGG human pathway and Reactome pathways enrichment analysis for 200 proteins identified by RWR algorithm using S-glycoprotein of HCoV-229E.



503

504 Figure 2. PPI-COEX multilayer analysis, based on human PPI interactome and Gene Coexpression network,  
505 with top 50 closest proteins/genes identified by RWR, using S-glycoprotein of HCoV-229E. Edges in blue  
506 represent PP interactions, while red edges are coexpressions.



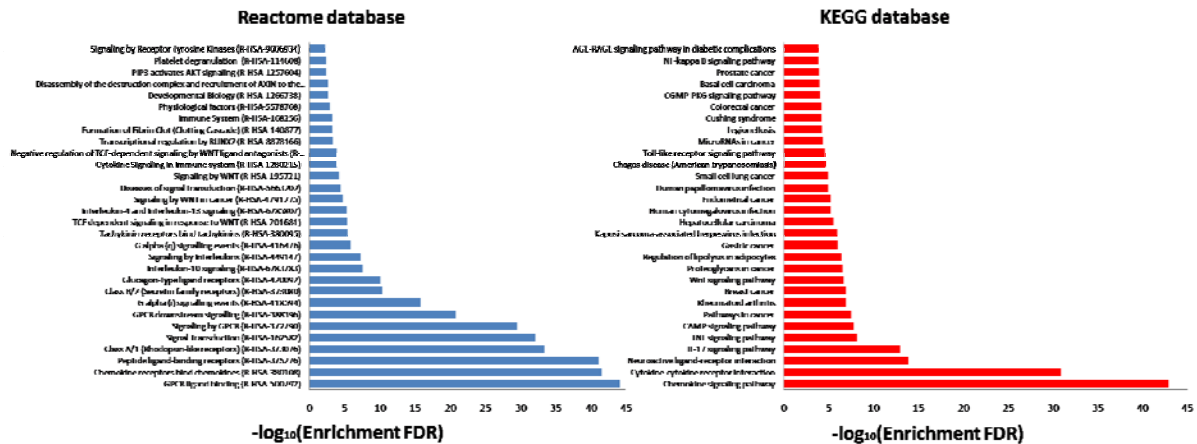
507

508 Figure 3. KEGG human pathway and Reactome pathways enrichment analyses for 200 proteins identified by

509 RWR algorithm using S-glycoprotein of SARS-CoV.



## Enrichment analyses for MERS-CoV



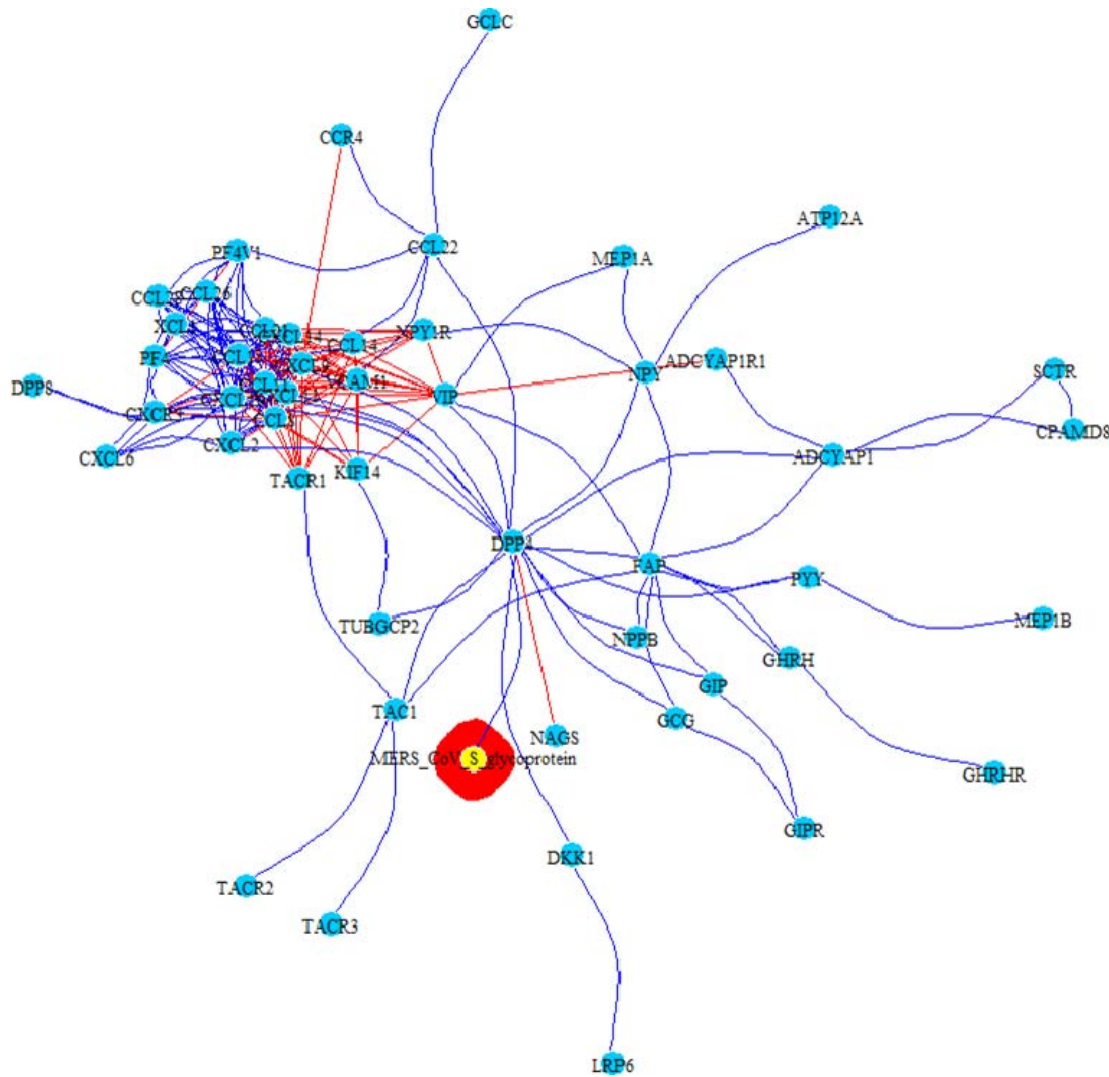
514

515 Figure 5. KEGG human pathway and Reactome pathways enrichment for 200 proteins identified by RWR

516 algorithm using S-Glycoprotein of MERS-CoV.

517





518

519 Figure 6. PPI-COEX multilayer analysis based on human PPI interactome and Gene Coexpression network,  
520 with top 50 closest proteins/genes identified by RWR, using S-glycoprotein of MERS-CoV. Edges in blue  
521 represent PP interactions, while red edges indicate coexpressions.

522

523

524 Figure S1. Pairwise distances along 259 full length CoV genomes. In the bottom of picture, indicative gene  
525 positioning along CoVs genomes is reported. The list of all considered genomes is reported in Table S1.

526

527 Figure S2. 3D structure of S-glycoprotein of SARS-CoV-2 and comparison with the ortholog from HCoV-  
528 229E, SARS-CoV, and MERS-CoV. Lateral (a) and superior (b) representation of SARS-CoV-2 S-  
529 glycoprotein, deduced for the sequence of patient INMI1 (MT066156.1). Each subunit chain has a different



530 color. Structure comparison of S-glycoprotein subunit between: HCoV-229E and SARS-CoV-2, in purple  
531 and blue respectively (c); SARS-CoV and SARS-CoV-2, in red and blue, respectively (d); MERS-CoV and  
532 SARS-CoV-2, in green and blue, respectively (e).

533

534 Figure S3. Amino acid alignment and secondary motifs in the receptor binding domain (RBD) of S-  
535 glycoprotein of HCoV-229E, SARS-CoV, MERS-CoV and SARS-CoV-2 are shown. Legend of secondary  
536 motifs identifiers: H =  $\alpha$  Helix, E =  $\beta$  Sheet, X = Random coil.

537

538 Figure S4. HCoV-229E - host interactome resulting from RWR applied to the top 200 closest proteins  
539 identified by RWR, using S-glycoprotein of HCoV-229E.

540

541 Figure S5. SARS-CoV - host interactome resulting from RWR applied to the top 200 closest proteins  
542 identified by RWR, using S-glycoprotein of SARS-CoV.

543

544 Figure S6. MERS-CoV - host interactome resulting from RWR applied to the top 200 closest proteins  
545 identified by RWR, using S-glycoprotein of MERS-CoV.

546

547 Table S1: List of accession numbers of H-CoV.

548

549 Table S2: list of genes selected by RWR algorithm for HCoV-229E, along with proximity score.

550

551 Table S3: list of genes selected by RWR algorithm for SARS-CoV, along with proximity score.

552

553 Table S4: list of genes selected by RWR algorithm for MERS-CoV, along with proximity score.

554