bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Analyzing the impact of SARS CoV-2 on the human proteome

Ernesto Estrada

Instituto Universitario de Matemáticas y Aplicaciones, Universidad de Zaragoza, 50009 Zaragoza, Spain; ARAID Foundation, Government of Aragón, 50018 Zaragoza, Spain.

Abstract

The COVID-19 respiratory disease is caused by the novel coronavirus SARS-CoV-2, which uses the enzyme ACE2 to entry human cells. This disease is characterized by important damages at multi-organ level, partially due to the abundant expression of ACE2 in practically all human tissues. However, not every organ in which ACE2 is abundant is affected by SARS CoV-2, which suggests the existence of other multi-organ routes for transmitting the perturbations produced by the virus. We consider here diffusive processes through the protein-protein interaction (PPI) network of proteins targeted by SARS CoV-2 as such alternative route. We found a subdiffusive regime that allows the propagation of virus perturbations through the PPI network at a significant rate. By following the main subdiffusive routes across the PPI network we identify proteins mainly expressed in the heart, cerebral cortex, thymus, testis, lymph node, kidney, among others of the organs reported to be affected by COVID-19.

1 Introduction

Since December 2019 a new coronavirus designated SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) (*1*) has produced an outbreak of pulmonary disease which has soon became a global pandemic (*2*, *3*). The new coronavirus disease (COVID-19) is characterized by a wide range of clinical manifestations (*4*, *5*), with an important implication of multi-organ and systemic dysfunctions (*6*, *7*), which include heart failure, renal failure, liver damage, shock and multi-organ failure. The new coronavirus shares about 82% of its genome with the one which produced the 2003 outbreak (SARS CoV) (*8*). Both coronaviruses share the same cellular receptor, which is the angiotensin-converting enzyme 2 (ACE2) receptor (*8*, *9*, *10*). ACE2 receptors are enriched in several tissues across the organs, such as in alveolar epithelial type II cells of lung tissues, in the heart, endothelium, kidneys and intestines. Therefore, ACE2 has been hypothesized as a potential cause of the major complications of the COVID-19 (*11*, *12*).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

However, it has been found that ACE2 has abundant expression on endothelia and smooth muscle cells of virtually all organs (*13*). Therefore, it should be expected that after SARS CoV-2 is present in circulation, it can be spread across all organs. In contrast, both SARS CoV and SARS CoV-2 are found specifically in some organs but not in others, as shown by *in situ* hybridization studies for SARS CoV [13]. This was already remarked by Hamming et al. (*13*) by stressing that it "*is remarkable that so few organs become viruspositive, despite the presence of ACE2 on the endothelia of all organs and SARS-CoV in blood plasma of infected individuals*".

Recently Gordon et al. (*14*) identified human proteins that interact physically with those of the SARS CoV-2 forming a high confidence SARS-CoV-2-human protein-protein interactions (PPIs) system. Using this information Gysi et al. (*15*) discovered that 208 of the human proteins targeted by SARS CoV-2 forms a connected component inside the human PPI network. That is, these 208 are not randomly distributed across the human proteome but they are closely interconnected by short routes that allow moving from one to another in just a few steps. These interdependencies of protein-protein interactions are known to enable that perturbations on one interaction propagates across the network and affect other interactions (*16*, *17*, *18*, *19*). In fact, it has been signified that diseases are a consequence of such perturbation propagation (*20*, *21*, *22*). It has been stressed that the protein-protein interaction process requires diffusion in their initial stages (*23*). The diffusive processes occur when proteins, possibly guided by electrostatic interactions, need to encounter each other many times before forming an intermediate (*24*). Not surprisingly, diffusive processes have guided several biologically oriented searches in PPI networks (*25*, *26*). Therefore, we assume here that perturbations produced by SARS CoV-2 proteins on human PPI network are propagated by means of diffusive processes. However, due to the crowded nature of the intra-cell space and the presence in it of spatial barriers, subdiffusive processes more than normal diffusion are expected for these protein-protein encounters (*27*, *28*, *29*). This creates another difficulty, as remarked by Batada et al. (*23*), which is that such (sub)diffusive processes along are not sufficient for carrying out cellular processes at a significant rate in cells.

Here we propose the use of a time-fractional diffusion model on the PPI network of proteins targeted by SARS CoV-2. The goal is to model the propagation of the perturbations produced by the interactions of human proteins with those of SARS CoV-2 through the whole PPI. The subdiffusive process emerging from the application of this model to the SARS-CoV-2-human PPIs has a very small rate of convergence to the steady state. However, this process produces dramatic increment of the probability that certain proteins are perturbed at very short times. This kind of shock wave effect of the transmission of perturbations occurs at much earlier times in the subdiffusive regime than at the normal diffusion one. Therefore, we propose here a switch and restart process in which a subdiffusive process starts at a given protein of the PPI, perturbs a few others, which then become the starting point of a new subdiffusive processm and so on. Using this approach we then analyze how the initial

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

interaction of the SARS CoV-2 Spike protein with a human protein propagates across the whole network. We discover some potential routes of propagation of these perturbations from proteins mainly expressed in the lungs to proteins mainly expressed in other different tissues, such as the heart, cerebral cortex, thymus, lymph node, testis, prostate, liver, small intestine, duodenum, kidney, among others.

2 Materials and Methods

2.1 Time-fractional diffusion model on networks

In this work we always consider $\Gamma = (V, E)$ to be an undirected finite network with vertices V representing proteins and edges E which represent the interaction between pairs of proteins. Let us consider $0 < \alpha \leq 1$ and a function $u : [0, \infty) \to \mathbb{R}$, then we denote by $D_t^\alpha u$ the fractional Caputo derivative of u of order α, which is given by (*31*)

$$D_t^\alpha u(t) = \int_0^t g_{1-\alpha}(t-\tau)\, u'(\tau)\, d\tau := (g_{1-\alpha} * u')(t), \quad t > 0,$$

where $*$ denotes the classical convolution product on $(0, \infty)$ and $g_\gamma(t) \coloneqq \frac{t^{\gamma-1}}{\Gamma(\gamma)}$, for $\gamma > 0$. Observe that the previous fractional derivative has sense whenever the function is derivable and the convolution is defined (for example if u' is locally integrable). The notation g_γ is very useful in the fractional calculus theory, mainly by the property $g_\gamma * g_\delta = g_{\gamma+\delta}$ for all $\gamma, \delta > 0$.

Here we propose to consider the time-fractional diffusion (TFD) equation on the network as

$$D_t^\alpha x(t) = -C\mathcal{L}x(t), \tag{1}$$

with initial condition $x(0) = x_0$, where $x_i(t)$ is the probability that protein i is perturbed at time t, C is the diffusion coefficient of the network, which we will set hereafter to unity, and $\mathcal{L}$ is the graph Laplacian, i.e., $\mathcal{L} = K - A$, where K is a diagonal matrix of node degrees and A is the adjacency matrix. This model was previously studied in distributed coordination algorithms for the consensus of multi-agent systems (*32*, *33*, *34*). The use of fractional calculus in the context of physical anomalous diffusion has been reviewed by Metzler and Klafter (*30*). The solution of this TFD equation is given by:

$$x(t) = E_{\alpha,1}(-t^\alpha \mathcal{L})\, x_0. \tag{2}$$

where $E_{\alpha,\beta}(\cdot)$ is the Mittag-Leffler matrix function. This result is proved in the Supplementary Note 1.

We can write $\mathcal{L} = U\Lambda U^{-1}$, where $U = \left[\vec{\psi}_1 \cdots \vec{\psi}_n\right]$ and $\Lambda = diag(\mu_r)$. Then,

$$E_{\alpha,1}(-t^\alpha \mathcal{L}) = U E_{\alpha,1}(-t^\alpha \Lambda) U^{-1}, \tag{3}$$

which can be expanded as

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$E_{\alpha,1}\left(-t^{\alpha}\mathcal{L}\right)=\vec{\psi}_1\vec{\phi}_1^{T}E_{\alpha,1}\left(-t^{\alpha}\mu_1\right)+\vec{\psi}_2\vec{\phi}_2^{T}E_{\alpha,1}\left(-t^{\alpha}\mu_2\right)+\cdots+\vec{\psi}_n\vec{\phi}_n^{T}E_{\alpha,1}\left(-t^{\alpha}\mu_n\right), \tag{4}$$

where $\vec{\psi}_j$ and $\vec{\phi}_j$ are the jth column of U and of U^{-1}, respectively. Because $\mu_1=0$ and $0<\mu_2\leq\cdots\leq\mu_n$ for a connected graph we have

$$\lim_{t\to\infty}E_{\alpha,1}\left(-t^{\alpha}\mathcal{L}\right)=\vec{\psi}_1\vec{\phi}_1^{T}, \tag{5}$$

where $\vec{\psi}_1^{T}\vec{\phi}_1=1$. Let us take $\vec{\psi}_1=\vec{1}$, such that we have

$$\lim_{t\to\infty}\vec{x}\left(t\right)=\lim_{t\to\infty}\left(E_{\alpha,1}\left(-t^{\alpha}\mathcal{L}\right)\right)\vec{x}_0=\left(\vec{1}\vec{\phi}_1^{T}\right)\vec{x}_0=\left(\vec{\phi}_1^{T}\vec{x}_0\right)\vec{1}. \tag{6}$$

This result indicates that in an undirected and connected network, the diffusive process controlled by the TFD equation always reaches a steady state which consists of the average of the values of the initial condition. Also, because the network is connected μ_2 makes the largest contribution to $E_{\alpha,1}\left(-t^{\alpha}\mathcal{L}\right)$ among all the nontrivial eigenvalues of $\mathcal{L}$. Therefore, it dictates the rate of convergence of the diffusion process. We remark that in practice the steady state $\lim_{t\to\infty}\left\|x_v\left(t\right)-x_w\left(t\right)\right\|=0,\forall v,w\in V$ is very difficult to achieve. Therefore, we use a threshold ε, e.g., $\varepsilon=10^{-3}$, such that $\lim_{t\to\infty}\left\|x_v\left(t\right)-x_w\left(t\right)\right\|=\varepsilon$ is achieved in relatively small simulation time.

Due to its importance in this work we remark the structural meaning of the Mittag-Leffler function of the Laplacian matrix appearing in the solution of the TFD equation. That is, $E_{\alpha,1}\left(-t^{\alpha}\mathcal{L}\right)$ is a matrix function, which is defined as

$$E_{\alpha,1}\left(-t^{\alpha}\mathcal{L}\right)=\sum_{k=0}^{\infty}\frac{\left(-t^{\alpha}\mathcal{L}\right)^{k}}{\Gamma\left(\alpha k+1\right)},\alpha>0, \tag{7}$$

where $\Gamma\left(\cdot\right)$ is the Euler gamma function. We remark that for $\alpha=1$ we recover the diffusion equation on the network: $dx\left(t\right)/dt=-\mathcal{L}x\left(t\right)$ and its solution $E_{1,1}\left(-t^{\alpha}\mathcal{L}\right)=\exp\left(-tC\mathcal{L}\right)$ is the well-known heat kernel of the graph.

2.2 Time-fractional diffusion distance

We define here a generalization of the diffusion distance studied by Coifman and Lafon (*35*). We start by defining

$$D_{vw}=f\left(\tau\mathcal{L}\right)_{vv}+f\left(\tau\mathcal{L}\right)_{ww}-2f\left(\tau\mathcal{L}\right)_{vw} \tag{8}$$

where $f\left(\tau\mathcal{L}\right)=E_{\alpha,1}\left(-t^{\alpha}\mathcal{L}\right)$. In the Supplementary Note 2 we prove that D_{vw} is a square Euclidean distance between the nodes v and w. Let μ_j be the jth eigenvalue and ψ_{ju} the uth entry of the jth eigenvector of the Laplacian matrix. Then, we can write the time-fractional diffusion distance as

$$D_{vw}=\sum_{j=1}^{n}\left(\psi_{ju}-\psi_{jv}\right)^{2}E_{\alpha,1}\left(-\tau\mu_j\right). \tag{9}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

It is evident that when $\alpha = 1$, D_{vw} is exactly the diffusion distance previously studied by Coifman and Lafon (*35*). The fractional-time diffusion distance between every pair of nodes in a network can be represented in a matrix form as follow:

$$\Sigma(\tau) = \left(\vec{s}(\tau)\,\vec{1}^T + \vec{1}\vec{s}(\tau)^T - 2f(\tau\mathcal{L})\right)^{\circ 1/2}, \tag{10}$$

where $\vec{s} = [f(\tau\mathcal{L})_{11}, f(\tau\mathcal{L})_{22}, \cdots, f(\tau\mathcal{L})_{nn}]$ is a vector whose entries are the main diagonal terms of the Mittag-Leffler matrix function, $\vec{1}$ is an all-ones vector and $\circ$ indicates an entrywise operation. Using this matrix we can build the diffusion distance-weighted adjacency matrix of the network:

$$W(\tau) = A \circ \Sigma(\tau) = \Sigma(\tau) \circ A. \tag{11}$$

The shortest diffusion path between two nodes is then the shortest weighted path in $W(\tau)$. Let us consider each of the terms forming the definition of the time-fractional diffusion distance and apply the limit of the very small $-t^\alpha$. That is,

$$\lim_{\tau\to 0} f(\tau\mathcal{L})_{vv} = \lim_{\tau\to 0} \sum_{r=1}^{n} \psi_{rv}^2 f(\tau\Lambda) = 1, \tag{12}$$

and in a similar way,

$$\begin{aligned} \lim_{\tau\to 0} f(\tau\mathcal{L})_{vw} &= \lim_{\tau\to 0} \sum_{r=1}^{n} \psi_{rv}\psi_{rw} f(\tau\Lambda) \\ &= \sum_{r=1}^{n} \psi_{rv}\psi_{rw} = 0. \end{aligned} \tag{13}$$

Therefore, $\lim_{\tau\to 0} D_{vw}(\tau) = 2$. Consequently, $\lim_{\tau\to 0} W(\tau) = 2A$, which immediately implies that the time-fractional shortest diffusion path is identical to the shortest (topological) one in the limit of very small $\tau = -t^\alpha$.

2.3 Network of proteins targeted by SARS CoV-2

The proteins of SARS CoV-2 and their interactions with human proteins were determined experimentally by Gordon et al. (*14*). Gysi et al. (*15*) constructed an interaction network of all 239 human proteins targeted by SARS CoV-2. In this network the nodes represent human proteins targeted by SARS CoV-2 and two nodes are connected if the corresponding proteins have been determined to interact with each other. Obviously, this network of proteins targeted by SARS CoV-2 is a subgraph of the protein-protein interaction (PPI) network of humans. One of the surprising findings of Gysi et al. (*15*) is the fact that this subgraph

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is not formed by proteins randomly distributed across the human PPI, but they form a main cluster of 208 proteins and a few small isolated components. Hereafter, we will always consider this connected component of human proteins targeted by SARS CoV-2. This network is formed by 193 proteins which are significantly expressed in the lungs. Gysi et al. (*15*) reported a protein as being significantly expressed in the lungs if its GTEx median value is larger than 5. GTEx (*36*) is a database containing the median gene expression from RNA-seq in different tissues. The other 15 proteins are mainly expressed in other tissues. However, in reporting here the tissues were proteins are mainly expressed we use the information reported in The Human Protein Atlas (*37*) were we use information not only from GTEx but also from HPA (see details at the Human Protein Atlas webpage) and FANTOM5 (*38*) datasets.

3 Results

3.1 Global characteristics of the fractional diffusion

The PPI network of human proteins targeted by SARS CoV-2 is very sparse, having 360 edges, i.e., its edge density is 0.0167, 30% of nodes have degree (number of connections per protein) equal to one and the maximum degree of a protein is 14. The second smallest eigenvalue of the Laplacian matrix of this network is very small, i.e., $\mu_2 = 0.0647$. Therefore, the rate of convergence to the steady state of the diffusion processes taking place on this PPI are very slow. We start by analyzing the effects of the fractional coefficient α on these diffusive dynamics. We use the normal diffusion $\alpha = 1$ as the reference system.

To analyze the effects of changing α over the diffusive dynamics on the PPI network we consider the solution of the TFD equation for processes starting at a protein with a large degree, i.e., PRKACA, degree 14, and a protein with a low degree, i.e., MRPS5, degree 3. That is, the initial condition vector consists of a vector having one at the entry corresponding to either PRKACA or MRPS5 ans zeroes elsewhere. In Fig. 1 we display the changes of the probability with the shortest path distance from the protein where the process starts. This distance corresponds to the number of steps that the perturbation needs to traverse to visit other proteins. For $\alpha = 1.0$ the shape of the curves in Fig. 1 are the characteristic ones for the Gaussian decay of the probability with distance. However, for $\alpha < 1$ we observe that such decay differs from that typical shape showing a faster initial decay followed by a slower one. In order to observe this effect in a better way we zoomed the region of distances from 2 to 4 (see 1 (b) and (d)). As can be seen for distances below 3 the curve for $\alpha = 1.0$ is on top of those for $\alpha < 1$ indicating a slower decay of the probability. After this distance, there is an inversion and the normal diffusion occurs at much faster rate than the other two for the longer distances. This is a characteristic signature of subdiffusive processes, which starts at much faster rates than a normal diffusive process and then continue at much slower rates. Therefore, here we observe that the subdiffusive dynamics are much faster at earlier times of the process, which is when the perturbation occurs to close nearest neighbors to the initial

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

point of perturbation.

To further investigate these characteristic effects of the subdiffusive dynamics we study the time evolution of a perturbation occurring at a given protein and its propagation across the whole PPI network. In Fig. 2 we illustrate these results for $\alpha = 1.0$ (a), $\alpha = 0.75$ (a), $\alpha = 0.5$ (c). As can be seen in the main plots of this Figure the rate of convergence of the processes to the steady state is much faster in the normal diffusion (a) than in the subdiffusive one (b) and (c). However, at very earlier times (see insets in Fig. 2) there is a shock wave increase of the perturbation at a set of nodes. Such kind of shock waves have been previously analyzed in other contexts as a way of propagating effects across PPI networks (*17*). The remarkable finding here is however the fact that such shock wave occurs at much earlier times in the subdiffusive regimes than at the normal diffusion. That is, while for $\alpha = 1.0$ these perturbations occur at $t \approx 0.1 - 0.3$, for $\alpha = 0.75$ they occur at $t \approx 0.0 - 0.2$, and for $\alpha = 0.5$ at $t \approx 0.0 - 0.1$. Seeing this phenomenon in the light of what we have observed in the previous paragraph is not strange due to the observation that such processes go at much faster rate at earlier times, and at short distances, than the normal diffusion. In fact, this is a consequence of the existence of a positive scalar T for which $E_{\alpha,1}\left(-\gamma t^{\alpha}\right)$ decreases faster than $\exp\left(-\gamma t\right)$ for $t \in (0, T)$ for $\gamma \in \mathbb{R}^{+}$ and $\alpha \in \mathbb{R}^{+}$ (see Theorem 4.1 in (*32*)). Hereafter, we will consider the value of $\alpha = 0.75$ for our experiments due to the fact that it reveals a subdiffusive regime, but the shoch waves observed before are not occurring in an almost instantaneous way like when $\alpha = 0.5$, which would be difficult from a biological perspective.

The previous results put us in a crossroad. First, the subdiffusive processes which are expected due to the crowded nature of the intra-cellular space, are very slow for carrying out cellular processes at a significant rate in cells. However, the perturbation shocks occurring at earlier times of these processes are significantly faster than in normal diffusion. To sort out these difficulties we propose a switching back and restart subdiffusive process occurring in the PPI network. That is, a subdiffusive process starts at a give protein, which is directly perturbed by a protein of SARS CoV-2. It produces a shock wave increase of the perturbation in close neighbors of that proteins. Then, a second subdiffusive process starts at these newly perturbed proteins, which will perturb their nearest neighbors. The process is repeated until the whole PPI network is perturbed. This kind of "switch and restart processes" have been proposed for engineering consensus protocols in multiagent systems (*40*) as a way to accelerate the algorithms using subdiffusive regimes.

3.2 Diffusion from SARS CoV-2 Spike protein

The so-called Spike protein (S-protein) of the SARS CoV-2 interacts with only two proteins in the human hosts, namely ZDHHC5 and GOLGA7. The first protein, ZDHHC5, is not in the main connected component of the PPI network of SARS CoV-2 targets. Therefore, we will consider here how a perturbation produced by the interaction of the virus S-protein with GOLGA7 is propagated

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

through the whole PPI network of SARS CoV-2 targets. GOLGA7 has degree one in this network and its diffusion is mainly to close neighbors, namely to proteins separated by 2-3 edges. When starting the diffusion process at the protein GOLGA7 the main increase in the probability of perturbing another protein is reached for protein GOLGA3, which increases its probability up to 0.15 at $t = 0.2$, followed by PRKAR2A, with a small increase in its probability, 0.0081. Then, the process switch and restarts at GOLGA3, which mainly triggers the probability of the protein PRKAR2A–a major hub of the network. Once we start the process at PRKAR2A, practically the whole network is perturbed with probabilities larger than 0.1 for 19 proteins apart from GOLGA3. These proteins are, in decreasing order of their probability of being perturbed: AKAP8, PRKAR2B, CEP350, MIB1, CDK5RAP2, CEP135, AKAP9, CEP250, PCNT, CEP43, PDE4DIP, PRKACA, TUB6CP3, TUB6CP2, CEP68, CLIP4, CNTRL, PLEKHA5, and NINL. Notice that the number of proteins perturbed is significantly larger than the degree of the activator, indicating that not only nearest neighbors are activated.

An important criterion for revealing the important role of the protein PRKAR2A as a main propagator in the network of proteins targeted by SARS CoV-2 is its average diffusion path length. This is the average number of steps that a diffusive process starting at this protein needs to perturb all the proteins in the network. We have calculated this number to be 3.6250, which is only slightly larger than the average (topological) path length, which is 3.5673. That is, in least than 4 steps the whole network of proteins is activated by a diffusive process starting at PRKAR2A. Also remarkable, the average shortest diffusive path length is almost identical to the shortest (topological) one. This means that this protein mainly uses shortest (topological) paths in perturbing other proteins in the PPI. In other words, it is highly efficient in conducting such perturbations. We will analyze this characteristics of the PPI of human proteins targeted by SARS CoV-2 in a further section of this work.

At this time almost any protein in the PPI network is already perturbed. Therefore we can switch and restart the subdiffusion from practically any protein at the PPI network. We then investigate which are the proteins with the higher capacity of activating other proteins which are involved in human diseases. Here we use the database SiGeNet (*39*), which is one of the largest publicly available collections of genes and variants associated to human diseases. We identified 38 proteins targeted by SARS CoV-2 for which there are "definitive" or "strong" evidence of being involved in a human disease or syndrome (see Supplementary Table 1). These proteins participate in 70 different human diseases or syndromes as given in Supplementary Tables 2 and 3. We performed an analysis in which a diffusive process starts at any protein of the network and we calculated the average probability that all the proteins involved in human diseases are then perturbed. For instance, for a subdiffusive process starting at the protein ARF6 we summed the probabilities that the 38 proteins involved in diseases are perturbed at an early time of the process $t = 0.2$. Then, we obtain a global perturbation probability of 0.874. By repeating this process for every protein as an initiator we obtained the top disease activators. We have

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

found that none of the 20 top activators is involved itself in any of the human diseases or syndromes considered here. They are, however, proteins which are important not because of their direct involvement in diseases or syndromes but because they propagate perturbations in a very effective way to those directly involved in such diseases/syndromes. Among the top activators we have found: ARF6, ECSIT, RETREG3, STOM, HDAC2, EXOSC5, THTPA, among others shown in Fig. 4, where we illustrate the PPI network of the proteins targeted by SARS CoV-2 remarking the top 20 disease activators.

3.3 Diffusion from the lungs to other organs

We now consider how a perturbation produced by SARS CoV-2 on a protein mainly expressed in the lungs can be propagated to proteins mainly located in other tissues (see Supplementary Table 4) by a subdiffusive process. That is, we start the subdiffusive process by perturbing a given protein which is mainly expressed in the lungs. Then, we observe the evolution of the perturbation at every of the proteins mainly expressed in other tissues. We repeat this process for all the 193 proteins mainly expressed in the lungs. In every case we record those proteins outside the lungs which are perturbed at very early times of the subdiffusive process. For instance, in Fig. 3 we illustrate one example in which the initiator is the protein GOLGA2, which triggers a shock wave on proteins RBM41, TL5 and PKP2, that are expressed mainly outside the lungs. We consider such perturbations only if they occur at $t < 1$.

Not every one of the proteins expressed outside the lungs is triggered by such shock waves at very early time of the diffusion. For instance, proteins MARK1 and SLC27A2are perturbed in very slow processes, and do not produce the characteristics high peaks in the probability at very short times. On the other hand, there are proteins expressed outside the lungs which are triggered by more than one protein from the lungs. The case of GOLGA2 is an example of a protein triggered by 3 ones in the lungs. In Table 1 we list some of the proteins expressed mainly in tissues outside the lungs which are heavily perturbed by proteins in the lungs. The complete list of the perturbing proteins is given in the Supplementary Table 5. We give three indicators of the importance of the perturbation of these proteins. They are: Act., which is the number of proteins in the lungs that activate each of them; p_{tot}, which is the sum of the probabilities of finding the diffusive particle at this protein for diffusive processes that have started in the their activators; t_{mean}, which is the average time required by activators to perturb the corresponding protein. For instance, PKP2 is perturbed by 21 proteins in the lungs, which indicates that this protein, mainly expressed in the heart muscle, has a large chance of being perturbed by diffusive processes starting in proteins mainly located at the lungs. Protein PRIM2 is activated by 5 proteins in the lungs, but if all these proteins were acting at the same time, the probability that PRIM2 is perturbed will be very high, $p_{\text{tot}} \approx 0.536$. Finally, protein TLE5 is perturbed by 13 proteins in the lungs, which needs as average $t_{\text{mean}} \approx 0.24$ to perturb TLE5. These proteins do not form a connected component among them in the network. The average shortest diffusion path

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between them is 5.286 with a maximum shortest subdiffusion path of 10. As average they are almost equidistant from the rest of the proteins in the network as among themselves. That is, the average shortest subdiffusion path between these proteins expressed outside the lungs and the rest of the proteins in the network is 5.106. Therefore, these proteins can be reached from other proteins outside the lungs in no more than 6 steps in subdiffusive processes like the ones considered here.

3.4 Subdiffusion paths

Finally we study here how the diffusive process determines the paths that the perturbation follows when diffusing from a protein to another not directly connected to it. The most efficient way of propagating a perturbation between the nodes of a network is through the shortest (topological) paths that connect them. The problem for a (sub)diffusive perturbation propagating between the nodes of a network is that it does not have complete information about the topology of the network as to know its shortest (topological) paths. The network formed by the proteins targeted by SARS CoV-2 is very sparse and this indeed facilitates that the perturbations occurs by following the shortest (topological) paths most of the time. Think for instance in a tree, which has the lowest possible edge density among all connected networks. In this case, the perturbation will always use the shortest (topological) paths connecting pairs of nodes. However, in the case of the PPI network studied here a normal diffusive process, i.e., $\alpha = 1$, not always uses the shortest (topological) paths. In this case, there are 1294 pairs of proteins for which the diffusive particle uses a shortest diffusive path which is one edge longer than the corresponding shortest (topological) path. This represents 6.11% of all total pairs of proteins that are interconnected by a path in the PPI network of proteins targeted by SARS CoV-2. However, when we have a subdiffusive process, i.e., $\alpha = 0.75$, this number is reduced to 437, which represents only 2.06% of all pairs of proteins. Therefore, the subdiffusion process studied here through the PPI network of proteins targeted by SARS CoV-2 has an efficiency of 97.9% relative to a process that always uses the shortest (topological) paths in hopping between proteins. In Fig. 5 we illustrate the frequency with which proteins not in the shortest (topological) paths are perturbed as a consequence that they are in the shortest subdiffusive paths between other proteins. For instance, the following is a shortest diffusive path between the two endpoints: RHOA-PRKACA-PRKAR2A-CEP43-RAB7A-ATP6AP1. The corresponding shortest (topological) path is: RHOA-MARK2-AP2M1-RAB7A-ATP6AP1, which is one edge smaller. The proteins PRKACA, PRKAR2A and CEP43 are those in the diffusive path which are not in the topological one. Repeating this selection process for all the diffusive paths that differs from the topological ones we obtained the results illustrated in Fig. 5. As can be seen there are 36 proteins visited by the shortest diffusive paths which are not visited by the corresponding topological ones. The average degree of these proteins is 7.28 and there is only a small positive trend between the degree of the proteins and the frequency with which they appear in these paths, e.g., the Pearson

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

correlation coefficient is 0.46.

4 Discussion

We have presented a methodology that allow the study of diffusive processes in (PPI) networks varying from normal to subdiffusive regimes. Here we have studied the particular case in which the time-fractional diffusion equation produces a subdiffusive regime, with the use of $\alpha = 3/4$ in the network of human proteins targeted by SARS CoV-2. A characteristic feature of this PPI network is that the second smallest eigenvalue is very small, i.e., $\mu_2 = 0.0647$. As this eigenvalue determines the rate of convergence to the steady state, the subdiffusive process converges very slowly to that state. What it has been surprising is that even in these conditions of very small convergence to the steady state, there is a very early increase of the probability in those proteins closely connected to the initiator of the diffusive process. That is, in a subdiffusive process on a network, the time at which a perturbation is transmitted from the initiator to any of its nearest neighbors occurs at an earlier time than for the normal diffusion. This is a consequence of the fact that $E_{\alpha,1}\left(-\gamma t^{\alpha}\right)$ decreases very fast at small values of t^{α}, which implies that the perturbation occurring at a protein i at $t = 0$ is transmitted almost instantaneously to the proteins closely connected to i. This effect may be responsible for the explanation about why subdiffusive processes, which are so globally slow, can carry out cellular processes at a significant rate in cells. We have considered here a mechanism consisting in switching and restarting several times during the global cellular process. For instance, a subdiffusive process starting at the protein i perturbs its nearest neighbors at very early times, among which we can find the protein j. Then, a new subdiffusive process can be restarted again at the node j, and so on.

One of the important findings of using the current model for the study of the PIN of proteins affected by SARS CoV-2 is the identification of those proteins which are expressed outside the lungs that can be more efficiently perturbed by those expressed in the lungs (see 1). For instance, the protein with the largest number of activators, PKP2, appears mainly in the heart muscle. It has been observed that the elevation of cardiac biomarkers is a prominent feature of COVID-19, which in general is associated with a worse prognosis (*41*). Myocardial damage and heart failure are responsible for 40% of death in the Wuhan cohort (see references in (*41*)). Although the exact mechanism involving the heart injury is not known the hypothesis of direct myocardial infection by SARS CoV-2 is a possibility, which act along or in combination with the increased cardiac stress due to respiratory failure and hypoxemia, and or with indirect injury from the systemic inflammatory response (*41*, *42*, *43*, *44*). As can be seen in n Table 1 testis is the tissue where several of the proteins targeted by SARS CoV-2 are mainly expressed, e.g., CEP43, TLE5, PRIM2, MIPOL1, REEP6, HOOK1, CENPF, TRIM59, and MARK1. Currently there is not conclusive evidence about testis damage by SARS CoV-2 (*45*, *46*, *47*, *48*). However, the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

previous SARS CoV that appeared in 2003 and which shares 82% of proteins with the current one, produced testis damage and spermatogenesis, and it was concluded that orchitis was a complication of that previous SARS disease (*45*). We also detect a few proteins mainly expressed in different brain tissues, such as CEP135, PRIM2, TRIM59, and MARK1. The implication of SARS CoV-2 and cerebrovascular diseases has been reported, including neurological manifestations as well as cerebrovascular disease, such as ischemic stroke, cerebral venous thrombosis, and cerebral hemorrhage (*49*, *50*, *51*).

Kidney damage in SARS CoV-2 patients has been reported (*52*, *53*, *54*), which includes signs of kidney dysfunctions, proteinuria, hematuria, increased levels of blood urea nitrogen, and increased levels of serum creatinine. As much as 25% of acute kidney injury has been reported in clinical setting of SARS CoV-2 patients. One of the potential mechanisms for kidney damage is the organ crosstalk (*52*), as can be the mechanism of diffusion from proteins in the lungs to proteins in the urinary tract and kidney proposed here. A very interesting observation from Table 1 is the existence of several proteins expressed mainly in the thymus and T-cells, such as: TLE5, RETREG3, RBM41, CENPF, and TRIM59. It has been reported that many of the patients affected by SARS CoV-2 in Wuhan displayed significant decrease of T-cells (*55*). Thymus is an organ which displays a progressive decline with age with reduction of the order of 3-5% a year until approximately 30-40 years of age and of about 1% per year after that age. Consequently, it was proposed that the role of thymus should be taken into account in order to explain why COVID-19 appears to be so mild in children (*55*). The protein TLE5 is also expressed significantly in the lymph nodes. It was found by Feng et al. (*56*) that SARS CoV-2 induces lymph follicle depletion, splenic nodule atrophy, histiocyte hyperplasia and lymphocyte reductions. The proteins HOOK1 and MIPOL1 are significantly expressed in the pituitary gland. There has been some evidence and concerns that COVID-19 may also damage the hypothalamo-pituitary-adrenal axis has been expressed by Pal (*57*), which may be connected with the participation of the before mentioned proteins.

Another surprising finding of the current work is the elevated number of subdiffusive shortest paths that coincide with the shortest (topological) paths connecting pairs of proteins in the PPI of human proteins targeted by SARS CoV-2. This means that the efficiency of the diffusive paths connecting pairs of nodes in this PPI is almost 98% in relation to a hypothetical process which uses the shortest (topological) paths in propagating perturbations between pairs of proteins. The 437 shortest diffusive paths reported here contain one more edge than the corresponding shortest (topological) paths. The proteins appearing in these paths would never be visited in the paths connecting two other proteins if only the shortest (topological) paths were used. What it is interesting is to note that 6 out of the 15 proteins which are mainly expressed outside the lungs are among the ones "crossed" by these paths. They are: TLE5 (thymus, lymph node, testis), PKP2 (heart muscle), CEP135 (skeletal muscle, heart muscle, cerebral cortex, cerebellum), CEP43 (testis), RBM41 (pancreas, T-cells, testis, retina) and RETREG3 (prostate, thymus). This means that the perturbation of these proteins occurs not only through the diffusion from other proteins in

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the lungs directly to them, but also through some "accidental" diffusive paths between pairs of proteins which are both located in the lungs.

All in all, the use of time-fractional diffusive models to study the propagation of perturbations on PPI networks seems a very promising approach. The model is not only biologically sounded but it also allows us to discover interesting hidden patterns of the interactions between proteins and the propagation of perturbations among them. In the case of the PIN of human proteins targeted by SARS CoV-2 our current finding may help to understand potential molecular mechanisms for the multi-organs and systemic failures occurring in many patients.

References

1. A. Gorbalenya, S. Baker, R. Baric, Coronaviridae study group of the International Committee on Taxonomy of Viruses: The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiol.* **5**, 536–544 (2020).

2. P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si,Y. Zhu, B. Li, C. L. Huang, H. D. Chen, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).

3. F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y Pei, M. L. Yuan, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020).

4. F. Jiang, L. Deng, L. Zhang, Y. Cai, C. W. Cheung, Z. Xia, Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). *J. Gen. Intern. Med.* **35**, 1545–1549 (2020).

5. B. Zhang, X. Zhou, Y. Qiu, F. Feng, J. Feng, Y. Jia, H. Zhu, K. Hu, J. Liu, Z. Liu, S. Wang, Y. Gong, Ch. Zhou, T. Zhu, Y. Cheng, Z. Liu, H. Deng, F. Tao, Y. Ren, B. Cheng, L. Gao, X. Wu, L. Yu, Z. Huang, Z. Mao, Q. Song, B. Zhu, J. Wang, Clinical characteristics of 82 death cases with COVID-19. medRxiv. (2020). doi: https://doi.org/10.1101/2020.02.26.20028191.

6. M. Zhou, X. Zhang, J. Qu, Coronavirus disease 2019 (COVID-19): a clinical update. *Front. Med.* **14**, 126–135 2020.

7. S. Zaim, J. H. Chong, V. Sankaranarayanan, A. Harky, COVID-19 and multi-organ response. *Curr. Probl. Cardiol.* (2020). https://doi.org/10.1016/j.cpcardiol.2020.100618 .

8. J. F. Chan, K. H. Kok, Z. Zhu, H. Chu, K. K. To, S. Yuan, K. Y. Yuen, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microb. Infect.* **9**, 221-236 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

9. W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, M. Farzan, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450–454 (2003).

10. M. Hoffmann,H. Kleine-Weber, N. Krueger, M. Müller, C. Drosten, S. Pöhlmann, The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. bioRxiv (2020). DOI: 10.1101/2020.01.31.929042.

11. X. Chai, L. Hu, Y. Zhang, W. Han, Z. Lu, A. Ke, J. Zhou, G. Shi, N. Fang, J. Fan, J. Cai, J. Fan, F. Lan, Specific ACE2 expression in cholangiocytes may cause liver damage after 2019-nCoV infection. bioRxiv (2020). DOI: 101101/20200203931766.

12. Y. Zhao, Z. Zhao, Y. Wang, Y. Zhou, Y. Ma, W. Zuo, Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. bioRxiv (2020). DOI: 101101/20200126919985.

13. I. Hamming, W. Timens, M. L. Bulthuis, A. T. Lely, G. J. Navis, H. van Goor, Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.* **203**, 631-637 (2004).

14. D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, M. J. O'meara, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Huttenhain, R. M. Kaake, A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing. *Nature* (2020). https://doi.org/10.1038/s41586-020-2286-9.

15. D. M. Gysi, I. D. Valle, M. Zitnik, A. Ameli, X. Gan, O. Varol, H. Sanchez, R. M. Baron, D. Ghiassian, J. Loscalzo, A. L. Barabási, Network medicine framework for identifying drug repurposing opportunities for covid-19. arXiv:2004.07229 (2020). https://arxiv.org/abs/2004.07229.

16. S. Maslov, K. Sneppen, Specificity and stability in topology of protein networks. *Science* **296**, 910-913 (2002).

17. M. A. Antal, C. Bode, P. Csermely, Perturbation waves in proteins and protein networks: applications of percolation and game theories in signaling and drug design. *Curr. Prot. Pept. Sci.* **10**, 161-72 (2009).

18. M. Santolini, A. L. Barabási, Predicting perturbation patterns from the topology of biological networks. *Proc. Natl. Acad. Sci.* **115**, E6375-83 (2018).

19. B. K. Stöcker, J. Köster, E. Zamir, S. Rahmann, Modeling and simulating networks of interdependent protein interactions. *Integrative Biol.* **10**, 290-305 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20. A. L. Barabási, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease. *Nature Rev. Genet.* **12**, 56-68 (2011).

21. I. W. Taylor, J. L. Wrana, Protein interaction networks in medicine and disease. *Proteomics* **12**, 1706-1716 (2012).

22. L. I. Furlong, Human diseases through the lens of network biology. *Trends Genet.* **29**, 150-159 (2013).

23. N. N. Batada, L. A. Shepp, D. O. Siegmund, Stochastic model of protein–protein interaction: Why signaling proteins need to be colocalized. *Proc. Natl. Acad. Sci.* **101**, 6445-6449 (2004).

24. M. Zacharias, Accounting for conformational changes during protein–protein docking. *Curr. Op. Struct. Biol.* **20**, 180-186 (2010).

25. L. Chen, Y. H. Zhang, Z. Zhang, T. Huang, Y. D. Cai, Inferring novel tumor suppressor genes with a protein-protein interaction network and network diffusion algorithms. *Mol. Therapy* **21**, 57-67 (2018).

26. A. Stojmirović, Y. K. Yu, Information flow in interaction networks. *J. Comput. Biol.* **14**, 1115-1143 (2007).

27. G. Guigas, M. Weiss, Sampling the cell with anomalous diffusion—the discovery of slowness. *Biophys. J.* **94**, 90-94 (2008).

28. S. Basak, S. Sengupta, K. Chattopadhyay, Understanding biochemical processes in the presence of sub-diffusive behavior of biomolecules in solution and living cells. *Biophys. Rev.* **11**, 1-22 (2019).

29. M. Woringer, X. Darzacq, Protein motion in the nucleus: from anomalous diffusion to weak interactions. *Biochem. Soc. Trans.* **46**, 945-956 (2018).

30. R. Metzler, J. Klafter, The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339**, 1-77 (2000).

31. F. Mainardi, *Fractional Calculus and Waves in Linear Viscoelasticity. An Introduction to Mathematical Models* (Imperial College Press, 2010).

32. Y. Cao, Y. Li, W. Ren, Y. Chen, Distributed coordination algorithms for multiple fractional-order systems. In *2008 47th IEEE Conference on Decision and Control* (IEEE, 2008) pp. 2920-2925.

33. Y. Cao, Y. Li, W. Ren, Y. Chen, Distributed coordination of networked fractional-order systems. *IEEE Trans. Syst. Man Cybern. B* **40**, 362-370 (2009).

34. J. Lu, J. Shen, J. Cao, J. Kurths, Consensus of networked multi-agent systems with delays and fractional-order dynamics. In Consensus and synchronization in complex networks 2013 (pp. 69-110). Springer, Berlin, Heidelberg.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

35. R. R. Coifman, S. Lafon, Diffusion maps. *Appl. Comput. Harm. Anal.* **21**, 5-30 (2006).

36. J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, B. Foster, The genotype-tissue expression (GTEx) project. *Nature Genetics* 45, 580-585 (2013).

37. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å Sivertsson, C. Kampf, E. Sjöstedt1, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. Al-Khalili Szigyarto, J. Odeberg, D. Djureinovic, J. Ottosson Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

38. The FANTOM Consortium and the RIKEN PMI and CLST (DGT), A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).

39. J. Piñero, J. M. Ramírez-Anguita, J. Saüch-Pitarch, F. Ronzano, E. Centeno, F. Sanz, L. I. Furlong, The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845-D55 (2020).

40. W. Sun, Y. Li, C. Li, Y. Chen, Convergence speed of a fractional order consensus algorithm over undirected scale-free networks. *Asian J. Control* **13**, 936-46(2011) .

41. A. Akhmerov, E. Marbán, COVID-19 and the heart, *Circulation Res.* **126**, 1443-1455 (2020).

42. R. M. Inciardi, L. Lupi, G. Zaccone, L. Italia, M. Raffo, D. Tomasoni, D. S. Cani, M. Cerini, D. Farina, E. Gavazzi, R. Maroldi, Cardiac involvement in a patient with coronavirus disease 2019 (COVID-19). *JAMA Cardiol.* (2020). doi:10.1001/jamacardio.2020.1096.

43. Y. Zheng, Y. Ma, J. Zhang, X. Xie, COVID-19 and the cardiovascular system. *Nat. Rev. Cardiol.* **17**, 259–260 (2020). https://doi.org/10.1038/s41569-020-0360-5.

44. K. J. Clerkin, J. A. Fried, J. Raikhelkar, G. Sayer, J. M. Griffin, A. Masoumi, S. S. Jain, D. Burkhoff, D. Kumaraiah, L. Rabbani, A. Schwartz, Coronavirus disease 2019 (COVID-19) and cardiovascular diseas, *Circulation* (2020). https://doi.org/10.1161/CIRCULATIONAHA.120.046941.

45. W. D. Maya, S. S. Du Plessis, P. A . Velilla, SARS-CoV-2 and the testis: similarity to other viruses and routes of infection. *Reproduct. BioMed. Online* (2020). DOI: https://doi.org/10.1016/j.rbmo.2020.04.009.

46. F. Chen, D. Lou, Rising concern on damaged testis of COVID-19 patients. *Urology* (2020). DOI:https://doi.org/10.1016/j.urology.2020.04.069.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

47. S. Wang, X. Zhou, T. Zhang, Z. Wang, The need for urogenital tract monitoring in COVID-19. *Nat. Rev. Urol.* (2020). https://doi.org/10.1038/s41585-020-0319-7.

48. C. Fan, K. Li, Y. Ding, W. L. Lu, J. Wang, ACE2 expression in kidney and testis may cause kidney and testis damage after 2019-nCoV infection. MedRxiv. (2020). doi: https://doi.org/10.1101/2020.02.12.20022418.

49. J. M. Saavedra, COVID-19, angiotensin receptor blockers, and the brain, *Cell. Mol. Neurobiol.* (2020). https://doi.org/10.1007/s10571-020-00861-y.

50. N. Lushina, J. S. Kuo, H. A. Shaikh, Pulmonary, cerebral, and renal thromboembolic disease associated with COVID-19 infection. *Radiology* (2020). https://doi.org/10.1148/radiol.2020201623.

51. B. Zhou, J. She, Y. Wang, X. Ma, A case of coronavirus disease 2019 with concomitant acute cerebral infarction and deep vein thrombosis. *Front. Neurol.* (2020). doi: 10.3389/fneur.2020.00296.

52. C. Ronco, T. Reis, Kidney involvement in COVID-19 and rationale for extracorporeal therapies. *Nature Rev. Nephrol.* (2020). https://doi.org/10.1038/s41581-020-0284-7.

53. V. Fanelli, M. Fiorentino, V. Cantaluppi, L. Gesualdo, G. Stallone, C. Ronco, G. Castellano, Acute kidney injury in SARS-CoV-2 infected patients. *Critical Care* (2020). https://doi.org/10.1186/s13054-020-02872-z.

54. Z. Li, M. Wu, J. Yao, J. Guo, X. Liao, S. Song, J. Li, G. Duan, Y. Zhou, X. Wu, Z. Zhou, Caution on kidney dysfunctions of COVID-19 patients, Available at SSRN: https://ssrn.com/abstract=3559601 or http://dx.doi.org/10.2139/ssrn.3559601.

55. A. Ruggiero, G. Attinà, A. Chiaretti, Additional hypotheses about why COVID-19 is milder in children than adults. *Acta Paediatrica* (2020). https://doi.org/10.1111/apa.15328.

56. Z. Feng, B. Diao, R. Wang, G. Wang, C. Wang, Y. Tan, L. Liu, C. Wang, Y. Liu, Y. Liu, Z. Yuan, The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) directly decimates human spleens and lymph nodes. medRxiv (2020). doi: https://doi.org/10.1101/2020.03.27.20045427.

57. R. Pal, COVID-19, hypothalamo-pituitary-adrenal axis and clinical implications, *Endocrine* (2020) https://doi.org/10.1007/s12020-020-02325-1.

Acknowledgements: The author thanks Dr. Deisy Morselli Gysi for sharing data and information. **Competing interests:** The author declares that he has no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and Tables

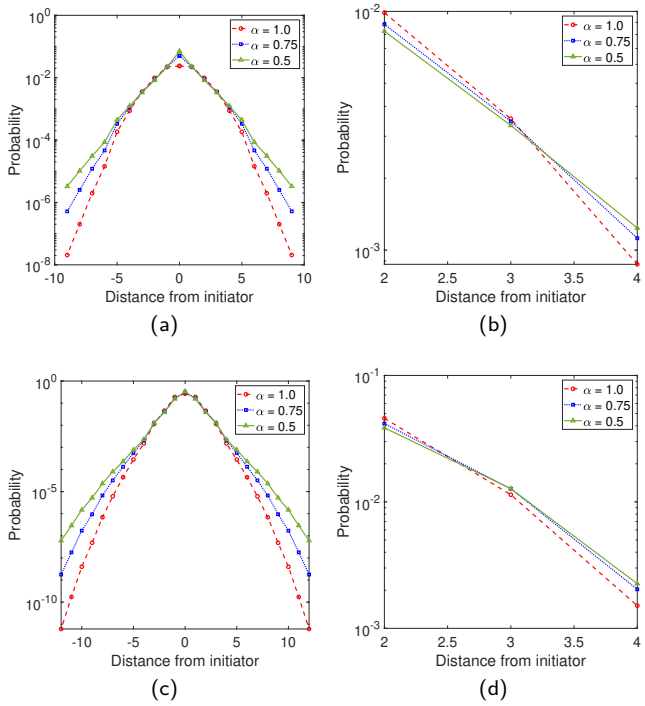

Fig. 1: **Spatial propagation of perturbations.** Average probability that a protein at a given distance from the perturbed protein feels the perturbation. (A) The initiators are the protein PRKACA (a) and (b) as well as MRPS5 (c) and (d). (b) and (d) zoom the scale of the processes in (A) and (C), respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

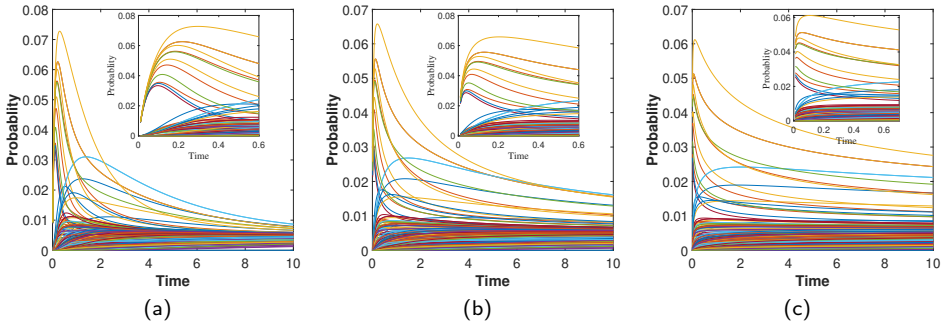

Fig. 2: **Shock wave increase of perturbations across the network**. Time evolution of the propagation of perturbations from the protein PRKACA to the rest of the proteins in the PPI network of human proteins targeted by SARS CoV-2. (a) Normal diffusion $\alpha = 1.0$. (b) Subdiffusion obtained for $\alpha = 3/4$. (c) Subdiffusion obtained for $\alpha = 1/2$. The insets illustrate the shortest time evolution of the perturbations.

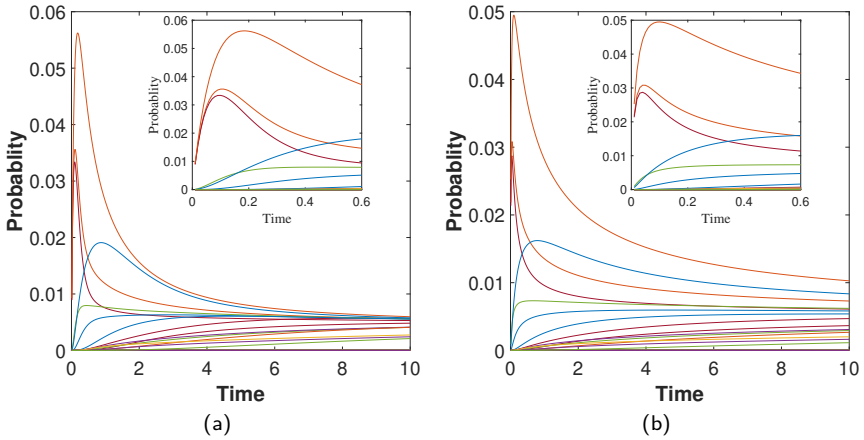

Fig. 3: **Shock wave increase of perturbations in proteins outside the lungs**.Time evolution of the propagation of perturbations from the protein GOLGA2 to proteins mainly expressed outside the lungs. (a) Normal diffusion $\alpha = 1.0$. (b) Subdiffusion with $\alpha = 3/4$. The insets illustrate the shortest time evolution of the perturbation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

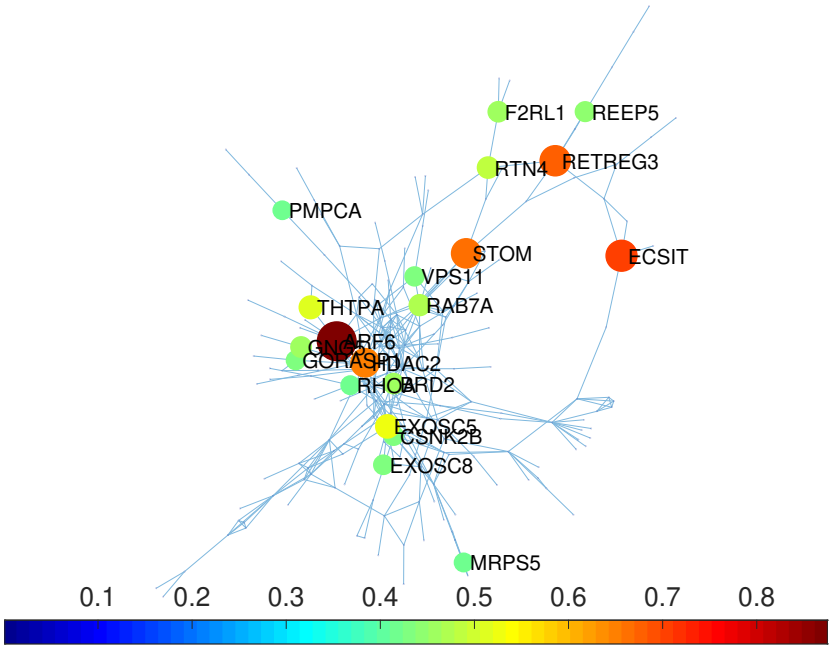

Fig. 4: **Main disease activators**. Proteins targeted by SARS CoV-2 identified as the top 20 main activators of proteins which are involved in human diseases.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

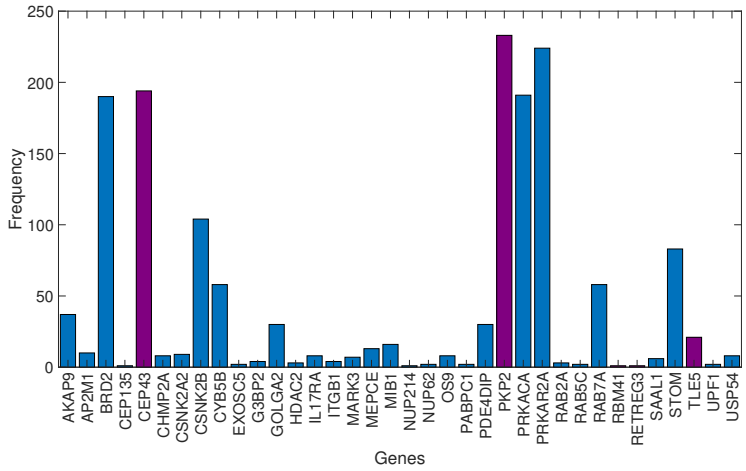

Fig. 5: **Proteins in shortest subdiffusive paths**. Frequency of proteins that appear in the shortest subdiffusion paths but not in the shortest (topological) paths connecting any pair of nodes in the PPI network. Bars marked in wine color are for those proteins expressed mainly outside the lungs.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

protein	Act.	p_{tot}	t_{mean}	tissues of main expression
PKP2	**21**	0.498	0.29	heart muscle
CEP43	18	0.521	0.26	testis
CEP135	14	0.527	0.30	skeletal muscle, heart muscle, cerebral cortex, cerebellum
TLE5	13	0.509	**0.24**	thymus, lymph node, testis
RETREG3	6	0.390	0.43	prostate, thymus
RBM41	6	0.209	0.49	pancreas, T-cells, testis, retina
PRIM2	5	**0.536**	0.30	cerebellum, parathyroid gland, testis
MIPOL1	3	0.155	0.57	pituitary gland, testis
REEP6	1	0.175	0.48	liver, small intestine, duodenum, testis
HOOK1	1	0.156	0.46	liver, parathyroid gland, testis, pituitary gland
CENPF	1	0.138	0.46	thymus, testis, bone marrow
ATP5ME	1	0.096	0.35	skeletal muscle, dendritic cells, heart muscle, brain
TRIM59	1	0.170	0.69	corpus callosum, brain, T-cells, testis
MARK1	1	N/A	>100	epididymis, cerebral cortex, heart muscle, testis, cerebellum
SLC27A2	1	N/A	>100	liver, kidney

Tab. 1: **Multi-organ propagation of perturbations**. Protein mainly expressed outside the lungs which are significantly perturbed during diffusive processes started at other proteins expressed in the lungs. Act., is the number of lungs proteins activators; p_{tot}, is the sum of the probabilities of finding the diffusive particle at this protein; t_{mean}, is the average time of activation (see text for explanations). The tissues of main expression are selected among the ones with the highest Consensus Normalized eXpression (NX) levels by combining the data from the three transcriptomics datasets (HPA, GTEx and FANTOM5) using the internal normalization pipeline (*37*).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5 Supplementary Information

Supplementary Note 1

We first prove that the solution of the FTD equation is given by $x(t) = E_{\alpha,1}(-t^{\alpha}\mathcal{L})\,x_0$, where $E_{\alpha,1}(M)$ is the Mittag-Leffler matrix function of M. We use the spectral decomposition of the network Laplacian $\mathcal{L} = U\Lambda U^{-1}$, where $U = \left[\vec{\psi}_1 \cdots \vec{\psi}_n\right]$ and $\Lambda = diag(\mu_r)$. Then we can write

$$D_t^{\alpha} x(t) = -U\Lambda U^{-1} x(t). \tag{14}$$

Let us define $y(t) = U^{-1}x(t)$, such that $D_t^{\alpha}x(t) = -U\Lambda y(t)$ and we have

$$\begin{aligned} U^{-1}D_t^{\alpha}x(t) &= -\Lambda y(t) \\ D_t^{\alpha}y(t) &= -\Lambda y(t). \end{aligned} \tag{15}$$

As Λ is a diagonal matrix we can write

$$D_t^{\alpha} y_i(t) = -\mu_i y_i(t), \tag{16}$$

which has the solution

$$y_i(t) = E_{\alpha,1}(-t^{\alpha}\mu_i)\, y_i(0). \tag{17}$$

We can replace $y_i(t) = U^{-1}x_i(t)$ to have

$$\begin{aligned} U^{-1}x_i(t) &= E_{\alpha,1}(-t^{\alpha}\mu_i)\, U^{-1}x_i(0) \\ x_i(t) &= UE_{\alpha,1}(-t^{\alpha}\mu_i)\, U^{-1}x_i(0), \end{aligned} \tag{18}$$

which finally gives the result in matrix-vector when written for all the nodes,

$$x(t) = E_{\alpha,1}(-t^{\alpha}\mathcal{L})\,x_0. \tag{19}$$

Supplementary Note 2

In this Note we prove that the function $D_{vw} = f(\tau\mathcal{L})_{vv} + f(\tau\mathcal{L})_{ww} - 2f(\tau\mathcal{L})_{vw}$, where $f(\tau\mathcal{L}) = E_{\alpha,1}(-t^{\alpha}\mathcal{L})$ is a Euclidean distance between the corresponding pair of nodes in the network. The matrix function $f(\tau\mathcal{L})$ can be written as $f(\tau\mathcal{L}) = Uf(\tau\Lambda)U^{-1}$. Let $\vec{\varphi}_u = [\psi_{1,u}, \psi_{2,u}, \cdots, \psi_{n,u}]^T$. Then,

$$D_{vw} = (\vec{\varphi}_v - \vec{\varphi}_w)^T f(\tau\Lambda)(\vec{\varphi}_v - \vec{\varphi}_w). \tag{20}$$

Therefore, because $f(\tau\mathcal{L})$ is positive defined we can write

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$
\begin{aligned}
D_{vw} &= \left(\left(\vec{\varphi}_v - \vec{\varphi}_w\right)^T f^{1/2}\left(\tau\Lambda\right)\right)\left(f^{1/2}\left(\tau\Lambda\right)\left(\vec{\varphi}_v - \vec{\varphi}_w\right)\right) \\
&= \left(f^{1/2}\left(\tau\Lambda\right)\vec{\varphi}_v - f^{1/2}\left(\tau\Lambda\right)\vec{\varphi}_w\right)^T\left(f^{1/2}\left(\tau\Lambda\right)\vec{\varphi}_v - f^{1/2}\left(\tau\Lambda\right)\vec{\varphi}_w\right) \\
&= \left(\vec{x}_v\left(\tau\right) - \vec{x}_w\left(\tau\right)\right)^T\left(\vec{x}_v\left(\tau\right) - \vec{x}_w\left(\tau\right)\right) \\
&= \left\|\vec{x}_v\left(\tau\right) - \vec{x}_w\left(\tau\right)\right\|^2,
\end{aligned}
\tag{21}
$$

where $\vec{x}_v\left(\tau\right) = f^{1/2}\left(\tau\Lambda\right)\vec{\varphi}_v$. Consequently, D_{vw} is a square Euclidean distance between v and w. In this sense the vector $\vec{x}_v\left(\tau\right) = f^{1/2}\left(\tau\Lambda\right)\vec{\varphi}_v$ is the position vector of the node v in the diffusion Euclidean space. Because $E_{1,1}\left(-t^\alpha\mathcal{L}\right) = \exp\left(-t^\alpha\mathcal{L}\right)$ we have that D_{vw} generalizes the diffusion distance studied by Coifman and Lafon, which is the particular case when $\alpha = 1$.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

protein	Disease ID
ACAD9	C1970173
AGPS	C0282529, C1838612
ALG8	C2931002, C4693472
AP2M1	C0004134, C0036572, C0557874, C0856975, C1858120, C3714756
AP3B1	C0079504, C0034069, C0024115
ATP6AP1	C4310819
ATP6V1A	C4479409, C4693934
BRD4	C0018798, C0025958, C3714756, C4025871
CDK5RAP2	C1858108, C0431350
CENPF	C0025958, C0152427, C0265202, C1292778
CEP135	C0431350
CIT	C4310723
CSDE1	C0004352, C0557874, C3714756
CYB5R3	C0025637
DNMT1	C3807295, C1384666
EXOSC3	C1843504, C0266468
GNB1	C0557874, C0036572, C3714756, C0424605
GPAA1	C0557874
IL17RA	C4310803, C3151402
LARP7	C3554439, C0013336, C0342573
MOGS	C1853736
MRPS2	C4693722
NDUFAF1	C4748769, C1838979, C0751651
NGLY1	C3808991
NSD2	C0025958, C0015934, C0026827, C3714756, C0456070, C4022738
PCNT	C0265202, C1292778
PKP2	C1836906, C0349788
PLOD2	C0432253, C0029434
PMPCB	C4693741
POLA1	C3860213
RALA	C0557874, C3714756, C0036572, C4022810
RDX	C3711374, C1970239, C1384666
REEP6	C4310626, C0035334
RPL36	C1260899
TAPT1	C0029434
TBK1	C0002736, C4225325, C4693542, C0751072
TCF12	C1856266
WFS1	C0043207, C3280358, C4518338, C0011860, C1384666, C0004138

Tab. 2: List of proteins targeted by SARS CoV-2 which are found in the database DisGeNet as displaying “definitive” or “strong” evidence of participating in human diseases. The Disease ID is the code of the disease in DisGeNet.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tab. 3: Diseases identified with "definitive" or "strong" evidence in humans in the database DisGeNet for proteins targeted by SARS CoV-2. The disease ID and names are from DisGeNet.

Disease ID	Disease/syndrome
C1970173	Acyl-CoA dehydrogenase family, Member 9, Ddeficiency of
C0282529	Chondrodysplasia Punctata, Rhizomelic
C1838612	Rhizomelic chondrodysplasia punctata, type 3
C2931002	Congenital disorder of glycosylation type 1H
C4693472	Polycystic liver disease 3 with or withouut kidney cysts
C0004134	Ataxia
C0036572	Seizures
C0557874	Global developmental delay
C0856975	Autistic behavior
C1858120	Generalized hypotonia
C3714756	Intellectual disability
C0079504	Hermanski-Pudlak syndrome
C0034069	Pulmonary fibrosis
C0024115	Lung diseases
C4310819	Immunodeficiency 47
C4479409	Cutis laxa, autosomal recessive, type IID
C4693934	Epileptic encephalopathy, infantile or early childhood, 3
C0018798	Congenital heart defects
C0025958	Microcephaly
C4025871	Abnormality of the face
C1858108	Microcephaly, primary autosomal recessive, 3
C0431350	Primary microcephaly
C0152427	Polydactyly
C0265202	Seckel syndrome
C1292778	Chronic myeloproliferative disorder
C4310723	Microcephaly 17, primary, autosomal recessive
C0004352	Autistic disorder
C0025637	Methemoglobinemia
C3807295	Cerebellar ataxia, deafness, and narcolepsy, autoomal dominant
C1384666	Hearing impairment
C1843504	Pontocerebellar Hhypoplasia Ttype 1
C0266468	Congenital pontocerebellar hypoplasia
C0424605	Developmental delay (disorder)
C4310803	Immunodeficiency 51
C3151402	Candidiasis, familial, 5
C3554439	Microcephalic primordial dwarfism Alazami type
C0013336	Dwarfism
C0342573	Pituitary dwarfism I
C1853736	Congenital disorder of glycosylation, Type IIB

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

C4693722	Combined oxydative phosporylation deficiency 36
C4748769	Mitochondrial complex I deficiency, nuclear type 11
C1838979	Mitochondrial complex I deficiency
C0751651	Mitochondrial diseases
C3808991	NGLY1 deficiency
C0015934	Fetal Ggrowth retardation
C0026827	Muscle hypotonia
C0456070	Growth delay
C4022738	Neurodevelopmental delay
C1836906	Arrhythmogenic right ventricular dysplasia, familial, 9
C0349788	Arrhythmogenic right Vventricular dysplasia
C0432253	Bruck syndrome
C0029434	Osteogenesis imperfecta
C4693741	Multiple mitochondrial dysfunctions syndrome 6
C3860213	Autoinflammatory disorder
C4022810	Abnormality of nervous system morphology
C3711374	Nonsyndromic deafness
C1970239	Deafness, autosomal recessive, 24
C4310626	Retinitis pigmentosa 77
C0035334	Retinitis Pigmentosa
C1260899	Anemia, Diamond-Blackfan
C0002736	Amyotrophic Lateral Sclerosis
C4225325	Frontotemporal dementia and/or amyotrophic lateral sclerosis 4
C4693542	Encephalopathy, acute, infection-induced (Herpes-specific), Susceptibility to, 8
C0751072	Frontotemporal lobar degeneration
C1856266	Coronal craniosynostosis
C0043207	Wolfram syndrome
C3280358	Wolfram-like syndrome, autosomal dominant
C4518338	Wolfram-like syndrome
C0011860	Diabetes Mmellitus, non-Insulin-dependent
C0004138	Ataxias, hereditary

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

protein	p_{tot}	protein	p_{tot}	protein	p_{tot}
ARF6	0.874	BRD2	0.4587	MRPS27	0.4099
ECSIT	0.6991	F2RL1	0.4575	EXOSC2	0.4065
RETREG3	0.6821	REEP5	0.4451	MARK3	0.3844
STOM	0.6614	EXOSC8	0.4354	PRIM1	0.3733
HDAC2	0.6421	CSNK2B	0.4334	POLA2	0.3733
EXOSC5	0.526	VPS11	0.429	PLEKHF2	0.3607
THTPA	0.5129	GORASP1	0.4244	RAB5C	0.3581
RTN4	0.4803	MRPS5	0.4185	G3BP1	0.3567
RAB7A	0.4724	RHOA	0.4181	PDE4DIP	0.3562
GNG5	0.4608	PMPCA	0.4143	TMEM97	0.3532

Tab. 4: Proteins with the largest effect on perturbing those 38 proteins identified in human diseases. p_{tot} is the sum of the probabilities that the given protein activates those identified as having "definitive" or "strong" evidence of being involved in a human disease.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

protein	consensus	HPA	GTEx	FANTOM5
PKP2	heart muscle	heart muscle fallopian tube	heart muscle	heart muscle
CEP43	testis	testis	testis	testis
CEP135	thymus	lymph node endometrium testis	testis	thymus
TLE5	skeletal muscle heart muscle	cerebral cortex	cerebellum	cerebellum heart muscle
RETREG3	pancreas	T-cells testis	tesitis	retina testis endometrium
RBM41	cerebellum parathyroid gland	parathyroid gland	testis thyroid gland	cerebellum
PRIM2	prostate thymus	prostate	N/A	thymus
MIPOL1	pituitary gland	testis	testis	pituitary gland
REEP6	liver small intestine	duodenum small intestine testis	testis	testis small intestine
HOOK1	liver	parathyroid gland	tesitis	pituitary gland kidney
CENPF	thymus	testis bone marrow	testis	thymus
ATP5ME	skeletal muscle	dendritic cells	heart muscle brain	heart muscle spinal cord mid brain
TRIM59	corpus callosum mid brain thalamus thymus	granullocites T-cells	testis spinal cord	thalamus mid brain corpus callosum
MARK1	epididymis cerebral cortex heart muscle	cerebral cortex testis epididymis	cerebellum	cerebellum
SLC27A2	liver	kidney liver	liver	liver

Tab. 5: RNA expression overview for proteins targeted by SARS CoV-2 and mainly expressed outside the lungs. We select the top RNA expressions in the four databases reported in The Human Protein Atlas.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.21.107912; this version posted May 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

protein	main perturbers from the lungs
PKP2	ARF6, RHOA, CSNK2A2, CSNK2B, GOLGA2, PRKACA, PRKAR2B, RAB1A, RAP1GDS1, BRD2, TLE1, NUP214, GPAA1, PMPCB, PDE4DIP, AKAP9, SUN2, NGDN, GRIPAP1, THTPA, ALG8
CEP43	PCNT, PRKAR2A, PRKAR2B, RAB7A, PDE4DIP, CEP350, AKAP9, TUBGCP3, TUBGCP2, CNTRL, CEP250, NINL, CEP68, PLEKHA5, CDK5RAP2, MIB1, GCC1, TIMM29
CEP135	GOLGB1, HDAC2, RAB8A, PRKAR2A, PRKAR2B, RAB1A, PDE4DIP, CEP350, ERC1, IL17RA, ZNF318, TBK1, PLEKHA5, CDK5RAP2, MIB1
TLE5	GOLGA2, TLE1, TLE3, EIF4E2, NUTF2, EXOSC8, IL17RA, EXOSC3, UBAP2, EXOSC5, GCC1, ZNF503, USP54, EXOSC2
RETREG3	STOM, REEP5, NDUFAF1, RTN4, PLEKHF2, CYB5B
RBM41	GOLGA2, RAB2A, TUBGCP2, NINL, GORASP1, GCC1
PRIM2	POLA1, PRIM1, RAE1, ERP44, POLA2
MIPOL1	EIF4E2, IL17RA, USP54
REEP6	REEP5
HOOK1	VPS39
CENPF	G3BP1
ATP5ME	SRP72
TRIM59	ECSIT
MARK1	MARK2
SLC27A2	USP13

Tab. 6: List of proteins mainly expressed outside the lungs and their major activators, which are proteins mainly expressed in the lungs.