

# Novel ACE2 Protein Interactions Relevant to COVID-19 Predicted by Evolutionary Rate Correlations

Austin Alves Varela<sup>1@</sup>, Sammy Cheng<sup>1</sup>, John Haynes Werren<sup>1@</sup>

<sup>1</sup> Department of Biology, University of Rochester, Rochester, NY, USA

@ Contributed equally

Corresponding Author:

John Werren<sup>1</sup>

402 Hutchinson Hall, Rochester, NY, 14620, USA

Email address: [jack.werren@rochester.edu](mailto:jack.werren@rochester.edu)

## Abstract

Angiotensin-converting enzyme 2 (ACE2) is the cell receptor that the coronavirus SARS-CoV-2 binds to and uses to enter and infect human cells. COVID-19, the pandemic disease caused by the coronavirus, involves diverse pathologies beyond those of a respiratory disease, including micro-thrombosis (micro-clotting), cytokine storms, and inflammatory responses affecting many organ systems. Longer-term chronic illness can persist for many months, often well after the pathogen is no longer detected. A better understanding of the proteins that ACE2 interacts with can reveal information relevant to these disease manifestations and possible avenues for treatment. We have undertaken an approach to predict candidate ACE2 interacting proteins which uses evolutionary inference to identify a set of mammalian proteins that “coevolve” with ACE2. The approach, called evolutionary rate correlation (ERC), detects proteins that show highly correlated evolutionary rates during mammalian evolution. Such proteins are candidates for biological interactions with the ACE2 receptor. The approach has uncovered a number of key ACE2 protein interactions of potential relevance to COVID-19 pathologies. Some proteins have previously been reported to be associated with severe COVID-19, but are not currently known to interact with ACE2, while additional predicted novel ACE2 interactors are of potential relevance to the disease. Using reciprocal rankings of protein ERCs, we have identified strongly interconnected ACE2 associated protein networks relevant to COVID-19 pathologies. ACE2 has clear connections to coagulation pathway proteins, such as Coagulation Factor V and fibrinogen components FGA, FGB, and FGG, the latter possibly mediated through ACE2 connections to Clusterin (which clears misfolded extracellular proteins) and GPR141 (whose functions are relatively unknown). ACE2 also connects to proteins involved in cytokine signaling and immune response (e.g. XCR1, IFNAR2, and TLR8), and to Androgen Receptor (AR). The ERC prescreening approach has elucidated possible functions for relatively uncharacterized proteins and possible new functions for well-characterized ones. Suggestions are made for the validation

of ERC-predicted ACE2 protein interactions. We propose that ACE2 has novel protein interactions that are disrupted during SARS-CoV-2 infection, contributing to the spectrum of COVID-19 pathologies.

## Introduction

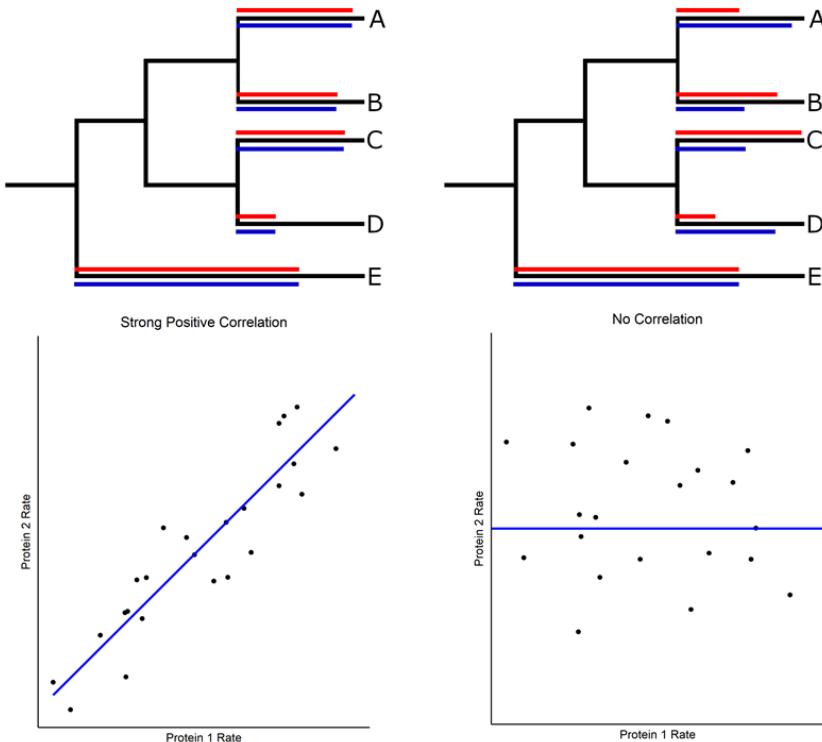
The coronavirus SARS-CoV-2 is causing severe pathologies and death among infected individuals across the planet. In addition to the symptoms expected from a respiratory disease, the infection can develop systemic manifestations (Gupta et al., 2020; Terpos et al., 2020; Siddiqi, Libby & Ridker, 2021). As a consequence, a wide range of pathologies are associated with COVID-19, including vascular system disruption, the extensive formation of blood clots (thrombosis) resulting in microvascular injury and stroke (Magro et al., 2020; Connors & Levy, 2020), gastrointestinal complications (Luo, Zhang & Xu, 2020) cardiac and kidney pathologies, ocular and dermatological symptoms (Bouaziz et al., 2020), neurological manifestations (Niazkar et al., 2020; Taquet et al., 2021), male infertility (Khalili et al., 2020), and a Kawasaki-like blood and heart disorder in children (Jones et al., 2020; Morand, Urbina & Fabre, 2020). A severe and often lethal immunoreaction can occur from respiratory and other infection sites, termed a “cytokine storm” (Chen et al., 2020). Even after acute SARS-CoV-2 infection has passed, individuals can suffer a suite of complications for many months, termed “Long Haul” syndrome (López-León et al., 2021), and the causes of these syndromes are not well understood.

Angiotensin-converting enzyme 2 (ACE2) is of obvious interest because it is a primary receptor for SARS-CoV-2 entry into human cells (Lan et al., 2020). However, ACE2 also plays a role in other important processes, such as regulation of blood pressure and vasodilation by the renin-angiotensin system (RAS), and protein digestion in the gut (Kuba et al., 2010). SARS-CoV-2 binding to ACE2 leads to a downregulation in ACE2 function (Verdecchia et al., 2020) which may be linked to the systemic damage by COVID-19 (Medina-Enríquez et al., 2020). It has been proposed that ACE2 receptor degradation during SARS-CoV-2 infection disrupts protection by ACE2 from inflammatory processes through the RAS and bradykinin pathways, possibly explaining patterns of COVID-19 severity with age and sex (Bastolla, 2020; Bastolla et al., 2021). As well as being a cell receptor, a circulating soluble form of the ectodomain of ACE2 (sACE2) is shed from cells and found in blood plasma, but the biological function of circulating ACE2 remains relatively unknown. Elevated levels of sACE2 have been detected in critically ill COVID-19 patients (van Lier et al., 2021) which coincides with a reduced expression of membrane-bound ACE2 (Medina-Enríquez et al., 2020), and a recent study indicates that sACE2 may assist SARS-CoV-2 entry into cells via other receptors (Yeung et al., 2021).

In general, ACE2’s protein-protein interaction network is likely to contribute to COVID-19 pathologies, due to ACE2’s role in systemic processes that are disrupted by the infection. Therefore, a fuller knowledge of ACE2 protein interactions is important to a better understanding of COVID-19 pathologies, including those that go beyond respiratory illness.

Common methods to identify protein-protein interactions include protein co-localization and precipitation, genetic manipulation, and proteomic profiling (Rao et al., 2014). Evolutionary approaches have also been used to evaluate protein interactions (De Juan, Pazos & Valencia, 2013), particularly to identify functional domains within proteins based on sequence conservation in evolution. Another set of methods utilize evolutionary rate correlations (also called evolutionary rate covariance or evolutionary rate coevolution). The concept is that coevolving proteins will show correlated rates of change across evolution (Wolfe & Clark, 2015). The approach has been used to detect physical interactions within and among proteins, as well as shared functionality not involving physical interaction, such as within metabolic pathways (Clark, Alani & Aquadro, 2012). For example, it has been employed to identify gene networks for post-mating response (Findlay et al., 2014), ubiquitination (Böhm et al., 2016), and recombination (Godin et al., 2015), and more recently to identify DNA repair genes (Brunette et al., 2019), cadherin-associated proteins (Raza et al., 2019), mitochondrial-nuclear interactions (Yan, Ye & Werren, 2019), and a mitochondrial associated zinc transporter (Kowalczyk et al., 2021), with subsequent experimental support. Evolutionary rate correlation (ERC) approaches are relatively inexpensive screening tools for detecting candidate protein interactions, and can also detect novel protein interactions that are not readily found in more traditional proteomic and genetic approaches (Colgren & Nichols, 2019; Yan, Ye & Werren, 2019). As such, “the ERC method should be a part of the toolkit of any experimental cell or developmental biologist” (Colgren & Nichols, 2019).

We have developed an evolutionary rate correlation (ERC) method that uses well-established phylogenies based on multiple lines of evidence (e.g. Misof et al. 2014 for insects and Kumar et al. 2017 for mammals) and calculates protein evolutionary rates for terminal branches for different proteins across a set of related species (Fig. 1). The approach is predicated on the idea that proteins that have strong evolutionary rate correlations are more likely to have functional interactions that are maintained by their coevolution, a conclusion supported by its predictive power in identifying known nuclear-mitochondrial encoded protein interactions in insects (Yan, Ye & Werren, 2019). That study also found that nuclear-encoded proteins and amino acids in contact with their mitochondrial-encoded components (e.g. oxidative phosphorylation proteins or mitochondrial ribosomal RNA) have significantly stronger ERCS than those not directly in contact. This result implicates physical interactions between proteins as one driver of evolutionary rate correlations, at least among nuclear-mitochondrial interactions. Other studies have found evolutionary rate correlations among proteins that do not make direct contact, such as in metabolic pathways (Clark, Alani & Aquadro, 2012).



**Figure 1: Evolutionary Rate Correlations.** The Spearman rank correlations between two proteins are calculated based on rates of protein evolution on terminal branches of a phylogeny. The relative rates of two proteins (red and blue lines) are shown in the hypothetical phylogenetic trees. Correlated and uncorrelated protein rates are illustrated below using a larger number of terminal branches (data points) than presented in the phylogeny.

We have developed a reciprocal rank approach to identify ACE2 associated networks and propose that these strongly coevolving proteins reveal ACE2 protein interactions that could be disrupted by COVID-19, thus contributing to its diverse pathologies. Particularly noteworthy are strong connections to coagulation pathway proteins, cytokine signaling, inflammation, immunity, and viral disease response.

It is important to note that our approach cannot be used to study coronavirus-ACE2 coevolution. The reason is that coronaviruses move between mammalian species and therefore do not have the same phylogenetic history as mammalian proteins, a prerequisite for the approach. We are also not asserting that coronavirus pressure is causing the evolution of ACE2 observed in the ERCs. Rather, it is our proposition that the ACE2 ERCs are revealing evolved mammalian protein

interactions that are not caused by, but could be relevant to COVID-19 pathologies, due to disruption of pathways in which ACE2 is normally involved.

We recognize that the predicted protein interactions detected by the ERC approach may not be causal for COVID-19 pathologies. However, the ERC connections to coagulation pathways, cytokine signaling, and immunity are striking and suggest a possible role of these ACE2 protein interactions in COVID-19 syndromes. The ERC results may also have implications for ACE2's role in the regulation of vasodilation through the renin-angiotensin system (RAS), cardiovascular disease, and protein digestion and absorption in the gut (Kuba et al 2010). Furthermore, the ERC analysis reveals possible novel connections for coagulation pathway and cytokine-signaling proteins that may be worthy of further investigation. Therefore, validation studies of the ERC predictions described here are desirable, both for possible applications to COVID-19 disease and treatment, and for investigations of other important biological processes.

## Materials & Methods

**Taxon Selection and Data Collection.** Our evolutionary rate correlation (ERC) approach requires orthologous protein sequence data across a large number of taxa with well-defined phylogenetic relationships. Calculation of evolutionary rates requires a resolved phylogeny of the taxa analyzed that is scaled to evolutionary time. Our ERC calculations utilize the TimeTree (Kumar et al., 2017) to generate a time-scaled phylogenetic tree using the mammalian taxa that are represented in OrthoDB sequence data (Supplementary Fig. S1). The tree generated is in units of millions of years and is based on a compilation of many phylogenetic-dating related studies. The tree is utilized as a base topology in phylogenetic analysis and its branch lengths are used to measure time for calculating evolutionary rates from the resultant individual protein trees (Fig. 1). Additional details on the data set are provided in the Supplementary Text, including Supplementary Figures and Tables. Large files are deposited in FigShare (available from the following link: <https://doi.org/10.6084/m9.figshare.14637450>) and listed in Supplementary Table S1.

Well-defined orthologous sequence data is sourced from OrthoDB (Kriventseva et al., 2019) at the “mammalia” (taxonomic id: 40674) taxonomic level. Since OrthoDB sequence data is gathered from a variety of sources and clustered algorithmically (unsupervised), primarily based on sequence similarity (Kriventseva et al., 2015), related paralogous proteins are often clustered with each other even if canonically annotated as functionally distinct proteins (Supplementary Table S2). Additionally, since the data sources for sequences can have varying levels of completeness, most ortholog groups on OrthoDB are missing sequence data for one or more taxa represented in the database. So, a majority of the data we initially selected was from single-copy ortholog groups with at least 90 of the 108 possible mammalian taxa present. In addition, some proteins with a possibly relevant function to COVID-19 pathologies (such as XCR1, and IFNAR2) or other relevant pathways in ortholog groups that did not meet the initial criteria, but that had minimal paralogy issues, were included. Paralogous sequences were manually

disambiguated based on published protein annotations and phylogenetic analysis. If a taxon in a given sequence had duplicate sequences that could not be disambiguated, the taxon was excluded in phylogenetic and ERC calculations for the specific proteins involved. In total, 1,953 orthologous protein groups are used in analyses.

**Phylogenetic Calculations and Protein Alignments.** To prepare orthologous sequence data for ERC calculation, each set of protein sequences are first aligned using the MAFFT software package (Katoh & Standley, 2013) using the following arguments: “--maxiterate 1000 --localpair --anysymbol”. Since the sequences come from data sources with varying levels of quality and multiple alignment programs can be imperfect, the aligned sequences must then be trimmed. The alignments are trimmed using the trimAl software package (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009) using the “-automated1” argument to remove poorly aligned regions. These final prepared alignments are then used to generate maximum-likelihood phylogenies. The IQ-TREE software package (Minh et al., 2020) is used to estimate protein branch lengths (equivalent to average substitution counts per site). Specifically, the “LG+F+G+I” model (which utilizes an empirically derived amino acid substitution matrix) is used with the following additional parameters: “-B 1000 -st AA -seed 1234567890” and the TimeTree phylogeny is provided to constrain output tree topology to reduce possible branch length estimation errors with the “-g” option. These trees are the basis of ERC calculations. Protein branch lengths are based on the average number of changes in amino acids at each residue in the alignment. The resultant branch lengths are paired with corresponding branches in the TimeTree to quantify branch-specific rates to be used for ERC calculations (described below). ERCs calculated with the more complete phylogeny (108 species) had short branch problems in oversampled taxonomic groups (described below and in the Supplementary Text). We therefore used a reduced phylogeny consisting of 60 taxa for subsequent ERC analyses.

**Calculation of ERCs.** Our evolutionary rate correlation (ERC) method is designed to predict protein-protein interactions using evolutionary data (Yan, Ye & Werren, 2019), and is based on protein evolutionary rates on terminal branches of the mammalian phylogeny (Fig. 1). We found that the more complete phylogeny (108 species) had short branch problems that inflate ERC spearman rank correlations (discussed in Supplementary Text). Most notably, there was an association between branch time and protein rate for many proteins, with oversampling in some taxonomic units (e.g. in Primates and Rodentia) leading to many ERCs being driven by relatively short branches (Supplementary Text). We attempted to control for these effects initially by using a partial correlation method, but found that it was not sufficient due to correlations between residuals and branch time (Supplementary Text). We then removed taxa that contributed short branches in our phylogeny based on either a 20MY or 30MY divergence time threshold (Supplementary Text) and recalculated branch rates for all proteins. We found that the 30MY threshold short branch removal eliminated significant branch time to protein rate correlations for the majority of proteins (87.5%). The resultant rate data no longer has branch time to branch rate as a confounding cofactor, and the ERCs themselves are no longer biased by extremely short

branches and taxonomic oversampling (Supplementary Text). The resulting data set is composed of 60 taxa and is used in our subsequent analyses of ERCs.

Using the adjusted data set, ERCs are calculated for every possible combination of protein pairs for which a tree has been generated. Every protein pair for which an ERC is calculated has each respective tree and the TimeTree topology is pruned to only include the shared taxa between the two, using the “ETE3” Python package (Huerta-Cepas, Serra & Bork, 2016). ETE3 is also used to extract the terminal branch lengths of each pruned tree. Evolutionary rates are calculated by dividing the terminal protein-tree branch lengths (average substitutions per site) by the corresponding branch in the TimeTree (measured in millions of years). Terminal branches are used for calculations as they do not have shared evolutionary histories and are therefore independent. The resulting rates have the unit of average substitution per site per millions of years. Given the resultant rates, evolutionary rate correlations are then calculated by performing a Spearman’s rank correlation test (Yan, Ye & Werren, 2019) using the Python package “SciPy” (Virtanen et al., 2020).

**Multiple Test Corrections.** P-values are corrected using the Benjamini-Hochberg FDR multiple-test correction procedure implemented in the Python package “statsmodels” (Seabold & Perktold, 2010). The FDR correction is applied to each respective protein’s set of ERCs. Correlation test results are non-directional, but FDR corrections are dependent on the rank of each correlation’s p-values. Since the rank of each correlation test value on respective protein lists vary, the FDR-corrected p-values of a given protein pair can differ depending on the specific ERC set. An ERC is considered significant if the FDR-corrected p-value is less than 0.05.

**ERC Set Enrichment Analysis.** To summarize the common biological function of proteins that tend to have strong ERCs, gene set enrichment analysis is performed on the top 2% of ERCs (by  $\rho$ ) of each protein(s) of interest (including the protein itself), including only proteins with ERCs that are significant following an FDR correction at a significance level of 0.05. At most, a protein of interest will have 41 proteins included for enrichment analysis (2% of the total 1953 proteins plus itself). Protein set enrichment analyses are performed using the Enrichr service (Xie et al., 2021) via the Python bindings provided by the “GSEAp” Python package (Fang et al., 2021) given the background of the full set of 1,953 proteins. We calculate enrichment results for ACE2 and all of its top 20 ERC partners. Additional enrichment analyses were also performed on a case-by-case basis based on relevance, including the reciprocal rank networks. Enrichments are performed using selected relevant term databases: KEGG\_2019\_Human, GO\_Biological\_Process\_2018, GO\_Cellular\_Component\_2018, GO\_Molecular\_Function\_2018, Reactome\_2016, WikiPathways\_2019\_Human, Tissue\_Protein\_Expr\_from\_Human\_Proteome\_Map, Tissue\_Protein\_Expr\_from\_ProteomicsDB, and Jensen\_TISSUES.

Enrichment results for terms that are significant at FDR-adjusted  $p < 0.05$  for all analyses are placed into a single table, organized by the enrichment term database (Supplementary File 3). The outputs from different databases can contain redundant terms to each other, so only the most significant of the redundant terms are reported for any enrichment analysis in the main text.

**Reciprocal Rank Network (RRN) Generation.** To evaluate and visualize the strongest ERCs centered around proteins of interest, “reciprocal rank networks” (RRNs) are produced. Reciprocal ranks refer to the fact that a significant ERC between two proteins can have different ranks in the two respective protein ERC lists because some proteins have more and higher ERCs than others. To focus on networks of proteins with strong reciprocal rank correlations, we have constructed networks based on proteins with reciprocal ranks of 20 or less (RR20), which is the top 1% in each protein's highest ERCs based on  $p$  values. Specifically, we have developed an ACE2 centric reciprocal rank network by the following steps (1) for ACE2, select its top 1% (20) proteins, (2) for each of those proteins, select additional proteins in their ERC list with reciprocal rank 20 or less, and then (3) Given the core set of proteins generated in the previous two steps, connect proteins which have a unidirectional rank of 20 or less.

The resultant network represents the strongest ERCs centered around a protein of interest (in this case ACE2), along with the immediate neighborhood of the strongest ERCs surrounding the protein of interest. The ACE2 Core Reciprocal Rank (**CRR**) was initiated with the four proteins to which ACE2 has RR20 ERCs (CLU, TMEM63C, FAM3D, and L1CAM), with GPR141 added due to its RR1 strong connection to CLU and unidirectional connection to ACE2. ACE2 also has highly significant ERCs to proteins that do not rank it in their top one percent.

Therefore, a similar approach has been used to generate an ACE2 reciprocal network initiated with the top 10 proteins to which ACE2 has highly significant ERCs, but are not reciprocally RR20 ranked, with a subsequent one cycle RR20 built upon these. This ACE2 Unidirectional Reciprocal Rank Network (**URR**) contains strong network connections to ACE2 through its high unidirectional ERCs. Steps 2-3 were omitted as the network becomes extremely large following just the first step, and our focus is on examining close connections to ACE2 based on ERC analysis.

**ERCs Within and Between Protein Complexes.** To compare whether calculated ERCs are stronger between known interactions versus non-interactions, the protein complex database, CORUM (Giurgiu et al., 2019), was used to retrieve known complexes. The “Core Complex” dataset was downloaded and filtered for human complexes to eliminate redundancy, resulting in 233 protein complexes from this CORUM data set which have two or more components present in our 1,953 protein ERC set, representing 258 pairwise ERC comparisons. As these protein complexes have redundancy (i.e. some complexes contain overlapping protein pairs), the set was further restricted to complexes containing unique protein components—resulting in 139 effective unique complexes considered. To test whether ERCs within complexes are higher than between complexes, all pairwise ERCs within complexes were compared to the median  $p$  value for each

pair to proteins present in non-redundant CORUM set that are not in complex with either of these proteins. A Wilcoxon matched signed-rank test was performed using the “wilcox.test” function in base R (version 3.6.1; with parameters “paired” and “exact” set to “TRUE”) on the in-complex  $\rho$  values and the median out-of-complex  $\rho$  values, to test if the in-complex  $\rho$  values were significantly greater than the median out-of-complex  $\rho$  values. In addition, as there were many complexes with a majority of subcomponents not present in our 1,953 datasets, the likelihood of individual pairs directly interacting within the complex decreases with the increasing number of proteins in a complex. Therefore, an additional Wilcoxon matched signed-rank test was performed on members of protein complexes composed of five or fewer proteins.

**Testing for Taxonomic Effects.** We use three methods to test for taxonomic effects on the calculated ERCs, (1) multiple linear regression, (2) analysis of covariance (ANCOVA), and (3) non-parametric independent contrasts (Garland, Harvey & Ives, 1992). For the regression and ANCOVA approaches, rate data was grouped by mammalian taxonomic orders accessed via ETE3 (Huerta-Cepas, Serra & Bork, 2016) and treated as an independent variable. The independent contrasts test uses the mammalian topology previously created with TimeTree (Kumar et al., 2017) to generate independent contrasts within the phylogeny. Statistical tests for each method are performed using base R (version 3.6.1). See the Supplementary Text for details.

**Testing Whether Branch Rates Increase When Extending Branch Time Within Clades.** To test whether increasing branch length results in increasing protein evolutionary rate, we selected separate phylogenetic groups (clades) from the full phylogeny (Supplementary Fig. S1) that contain short branch lengths. Protein evolutionary rate was calculated for each protein on the short branch, and then sequentially recalculated after removing adjacent taxa to extend the branch internally (Supplementary Fig. S5). In this way, the protein evolutionary rate was examined as branches are extended internally in independent clades within the tree. Comparing original branches to the 20MY correction resulted in 12 clades for which time scales change between 20MY and 30MY corrections, and 16 clades for which time scales change between 0MY and 30MY. Tests on each branch’s rate against the respective adjusted rate were performed using two-tailed Wilcoxon Matched Signed Rank Tests (Base R v3.6.1), first for proteins of interest (e.g. ACE2) and then for the full protein set. Results are described in the Supplementary Text.

## Results

**A. Basic Approach.** The basic methods are outlined here to provide context for the results which follow. To identify candidate protein interactions using evolutionary rate correlation, we utilized the consensus TimeTree phylogenetic reconstruction for mammalian species (Kumar et al., 2017). A total of 1,953 proteins (including ACE2) were aligned and evolutionary rates for each protein were then calculated for terminal branches of the tree (Fig. 1). This was determined by dividing the protein-specific branch length on each terminal branch by terminal branch time from the consensus tree (Yan, Ye & Werren, 2019). Maximum likelihood branch lengths were

estimated in IQ-TREE (Minh et al., 2020) using an empirical amino acid substitution matrix (see methods for details). To investigate evolutionary rate correlations (ERCs) among proteins, Spearman rank correlations were calculated for every protein pair using terminal branch rates (Fig. 1). Due to the large number of comparisons, a Benjamini-Hochberg false discovery rate (FDR) correction was calculated for each protein's ERC set (significance threshold  $\alpha = 0.05$ ). We subsequently found that many proteins show a positive correlation between terminal branch time and evolutionary rate, and observed that short branches in relatively oversampled taxa significantly contributed to this correlation (Supplementary Text). We, therefore, removed species that accounted for short branches, which eliminated the protein evolutionary rate to branch time correlation (see Methods and Supplementary Text for details). ERCs were then recalculated, and our ERC analyses are based on this set of 60 taxa.

In addition, we tested whether the observed lower rates of protein evolution for short terminal branches in the phylogeny are due to rates actually increasing over evolutionary time, versus a taxonomic effect. This was accomplished by examining changes in protein rates in independent clades as terminal branches were effectively extended by selective removal of flanking taxa. The analysis shows that evolutionary rates for many proteins increase as branch length is increased (described in more detail in Supplementary Text, Supplementary Fig. S4). A possible explanation for the pattern is that protein coevolution is mostly episodic, and short branches in a phylogeny are less likely to capture such events. In additional analyses, we tested for but did not find significant confounding effects of taxonomy on the ERC results (Supplementary Text).

Our analyses are focused on candidate protein interactions involving ACE2 using evidence of highly significant ERCs. For this purpose, we first examine proteins in ACE2's highest 2% of ERCs (top 40 proteins), all of which are highly significant after FDR correction (Table 1). Some of these ACE2 ERC proteins have been previously implicated in severe COVID-19 or SARS-CoV-2 gene expression effects on infected cells. However, while they have not been previously identified as having protein interactions with ACE2, this is predicted by our ERC analysis.

ACE2's						GEN1's					
Protein	ACE2 Rank	Partner Rank	<i>p</i>	P	FDR	Protein	GEN1 Rank	Partner Rank	<i>p</i>	P	FDR
GEN1	1	203	0.67	4.3E-08	4.2E-05	<b>IFNLR1**</b>	1	1	0.89	3.2E-20	6.2E-17
XCR1	2	37	0.67	3.2E-08	4.2E-05	<b>CC2D1B**</b>	2	1	0.84	5.3E-16	5.2E-13
<b>CLU**</b>	3	8	0.63	3.1E-07	1.5E-04	<b>MUC15**</b>	3	15	0.84	4.2E-15	2.7E-12
<b>TMEM63C**</b>	4	11	0.63	2.0E-07	1.3E-04	SPZ1	4	30	0.82	5.0E-14	1.4E-11
IFNAR2	5	392	0.62	2.5E-06	6.1E-04	<b>SLC10A6**</b>	5	2	0.82	1.2E-14	5.9E-12
KIF3B	6	26	0.60	1.7E-06	4.9E-04	<b>ARID4A**</b>	6	9	0.81	2.0E-14	8.0E-12
ITPR1PL2	7	364	0.59	1.7E-06	4.9E-04	RAD51AP2	7	22	0.81	6.7E-14	1.6E-11
FAM227A	8	175	0.59	1.8E-06	4.9E-04	<b>TESPA1**</b>	8	2	0.81	3.9E-14	1.3E-11
TLR8	9	243	0.58	3.7E-06	7.2E-04	<b>IFNAR2**</b>	9	9	0.80	3.4E-12	2.6E-10
COL4A4	10	541	0.58	3.7E-06	7.2E-04	<b>BCL6B**</b>	10	1	0.80	1.6E-13	3.6E-11
<b>FAM3D**</b>	11	2	0.57	5.8E-06	8.4E-04	RTL9	11	54	0.80	8.0E-13	1.1E-10
F5	12	642	0.57	4.1E-06	7.2E-04	<b>COL4A5**</b>	12	8	0.80	4.9E-13	8.7E-11
AR	13	22	0.57	7.7E-06	8.8E-04	APOBR	13	72	0.80	1.2E-12	1.3E-10
TSGA13	14	423	0.57	7.1E-06	8.8E-04	<b>COL4A6**</b>	14	19	0.79	1.6E-12	1.6E-10
PLA2G7	15	387	0.57	6.0E-06	8.4E-04	<b>TRADD**</b>	15	6	0.79	6.8E-13	1.0E-10
MMS19	16	387	0.56	5.9E-06	8.4E-04	FANCG	16	69	0.79	4.2E-13	8.2E-11
AMOT	17	124	0.56	8.1E-06	8.8E-04	CD180	17	27	0.78	8.4E-13	1.1E-10
<b>L1CAM**</b>	18	14	0.56	8.6E-06	8.8E-04	<b>TNFSF18**</b>	18	7	0.78	2.6E-12	2.2E-10
PDYN	19	428	0.56	7.3E-06	8.8E-04	<b>APOB**</b>	19	1	0.78	6.7E-13	1.0E-10
IQCD	20	158	0.56	9.2E-06	8.9E-04	<b>MKKS**</b>	20	20	0.78	8.7E-13	1.1E-10
SERPINA5	21	468	0.56	2.2E-05	1.4E-03	PIGV	21	8	0.78	1.6E-12	1.6E-10
CERS4	22	67	0.55	2.9E-05	1.5E-03	CCDC17	22	30	0.78	1.2E-12	1.3E-10
CC2D1B	23	467	0.55	1.1E-05	1.0E-03	DYTN	23	42	0.78	8.3E-12	5.1E-10
GPR141	24	17	0.55	1.5E-05	1.2E-03	GNPTAB	24	36	0.77	1.7E-12	1.6E-10
FSCB	25	817	0.55	2.8E-05	1.5E-03	MTMR11	25	13	0.77	2.9E-12	2.3E-10
RGR	26	167	0.55	3.0E-05	1.5E-03	TNFRSF1A	26	25	0.77	2.0E-12	1.7E-10
COL4A5	27	529	0.55	2.1E-05	1.4E-03	IFNAR1	27	5	0.77	2.7E-11	1.4E-09
TNFSF8	28	410	0.55	1.2E-05	1.1E-03	F2RL2	28	5	0.77	1.9E-11	1.1E-09
CCDC36	29	576	0.55	1.5E-05	1.2E-03	CXCR6	29	1	0.77	3.1E-11	1.5E-09
MRC1	30	195	0.55	1.3E-05	1.1E-03	KLHL6	30	6	0.77	3.3E-12	2.6E-10
CD27	31	550	0.54	3.0E-05	1.5E-03	SERPINA5	31	12	0.77	2.0E-11	1.1E-09
ADCK4	32	28	0.54	2.1E-05	1.4E-03	PLA2R1	32	31	0.77	6.6E-12	4.6E-10
SOWAHA	33	154	0.54	2.2E-05	1.4E-03	MYCBPAP	33	3	0.76	4.5E-12	3.3E-10
F2RL2	34	436	0.54	3.7E-05	1.7E-03	BPIFB2	34	5	0.76	7.6E-12	4.8E-10
WDR66	35	302	0.54	2.1E-05	1.4E-03	TLR7	35	114	0.76	1.4E-11	8.3E-10
TRADD	36	596	0.54	2.6E-05	1.5E-03	CCDC190	36	19	0.76	2.4E-10	6.2E-09
RELA	37	70	0.53	2.8E-05	1.5E-03	KMT2D	37	95	0.76	7.1E-12	4.8E-10
SLC10A6	38	533	0.53	3.0E-05	1.5E-03	FSCB	38	130	0.76	6.5E-11	2.6E-09
IL23A	39	383	0.53	4.7E-05	1.7E-03	CD27	39	19	0.76	2.7E-11	1.4E-09
TNFSF18	40	656	0.53	5.8E-05	1.8E-03	SNX11	40	24	0.76	7.3E-12	4.8E-10

**Table 1: Top 2% ERCs for ACE2 and GEN1.** The top two percent (2%) of ERCs are shown for ACE2 and GEN1, ranked by descending *p* value. The table illustrates how reciprocal ranks can differ between proteins with significant evolutionary correlations, depending on how interconnected proteins are. GEN1 has many partners which rank GEN1 highly in their respective ERCs. Also indicated in the table are examples of reciprocal rank correlations in which both partners rank the other in their top 20 (indicated by bold and asterisks). These are used to construct reciprocal rank protein interaction networks.

X-C Motif Chemokine Receptor 1 (XCR1) provides an illustrative example. XCR1 is a cytokine signaling receptor and ACE2's 2nd highest ranked ERC, with a highly significant evolutionary rate correlation. XCR1 is in a small genomic region that is implicated in severe COVID-19 by genome-wide association studies (Severe Covid-19 GWAS Group, 2020; Fricke-Galindo &

Falfán-Valencia, 2021). Another example is Interferon alpha/beta receptor 2 (IFNAR2) which, in a genome-wide association study (GWAS) and multi-omic analysis by Pairo-Castineira et al. (2021), was implicated in severe COVID-19. We therefore added it to our analysis, and surprisingly found it to be highly ranked (5th) among ACE2 ERCs. Clusterin (CLU) is the 3rd strongest ERC of ACE2 and the ACE2-CLU pair show high reciprocal ranks to each other (3rd in ACE2's set, 8th in CLU's set). CLU prevents the aggregation of misfolded proteins in the blood and delivers them to cells for degradation in lysosomes (Sánchez-Martín & Komatsu, 2020). CLU connects to key proteins in the coagulation pathway based on its reciprocal rank network (Section C, Fig. 2). CLU has been implicated in coronavirus infections, as one of only two proteins showing significant expression changes in cells infected by three different coronaviruses tested, including SARS-CoV-2 (Singh et al., 2021). The examples above lend credence to the proposition that the ERC approach is detecting ACE2 protein interactions that have implications to COVID-19.

Differences in ERC rank between protein pairs for the same correlation can occur because some proteins have higher and more extensive ERC connections than others. As a result, while two proteins can have a significant ERC with each other, each one's rank may differ in their respective ERC lists, as illustrated for ACE2 and GEN1 (Table 1). GEN1 (Flap endonuclease GEN homolog 1) is ACE2's top-ranked ERC, and is a DNA nuclease whose primary functions are resolution of DNA Holliday junctions and DNA damage checkpoint signaling (Chan & West, 2015). This protein shows high ERCs and is ranked highly in the ERC sets for many other proteins, suggesting central connectivity. As described further in Section C2, GEN1 shows unexpected enrichments for immune functions, perhaps related to its role in DNA damage checkpoint signaling.

Because our focus is on identifying strong candidate interactions involving ACE2 and its predicted partners, we utilize the rank information to identify proteins with high reciprocal ranks. Specifically, we focus on the strongest reciprocal ranks (RR) defined by ranks of less than or equal to 20 (RR20), which is the highest one percent of each protein's ERCs, and use these to develop reciprocal rank networks (Section C). Although speculative, we posit that protein pairs with high reciprocal ranks are likely to be strongly coevolving (i.e. both partners evolving reciprocally due to selective pressures acting on interacting domains between them). In contrast, protein pairs with a significant evolutionary rate correlation only one ranks highly (e.g. within the top two percent) in the ERC set of the other, are more likely to be due to "unidirectional" evolution. The rationale is that proteins with many significant ERC partners are under selective pressures primarily from their top evolutionary partners, whereas other interactors evolve primarily in response to the forces shaped by their stronger partner(s). We emphasize that this interpretation is speculative, and requires further exploration to determine what factors shape reciprocal ERC ranks between proteins.

The view that ERCs are detecting protein interactions relevant to COVID-19 is further supported by the analysis of ACE2 reciprocal rank ERC networks (Section C). Noteworthy in this regard are additional proteins in the coagulation pathway, such as Coagulation Factor V (F5), Fibrinogen Alpha Chain (FGA), Fibrinogen Beta Chain (FGB), and Fibrinogen Gamma Chain (FGG). Thrombosis (blood clotting) is a major pathology of COVID-19 (Gupta et al., 2020). Connections of ACE2 with the proteins above could relate to severe blood clotting problems in COVID-19 infections. ACE2 networks also show strong enrichments of cytokine signaling, viral (and pathogen) infections, and inflammatory response terms (Supplementary File 3), which are clearly relevant to COVID-19 pathologies such as cytokine storms and systemic inflammation.

In yet other cases, we have found proteins with significant ACE2 ERCs or ACE2 network connections, but for which there is little functional information, such as GPR141. We can use their ERCs to suggest possible functions for future investigation. Finally, ERCs for proteins of known function (such as F5 and GEN1) indicate likely additional roles, suggesting these proteins have unrecognized “moonlighting” functions (Jeffery, 1999).

Below, we first describe proteins of interest to which ACE2 has significant ERCs, summarize aspects of their known biological functions, and examine significantly enriched functional categories for these ERCs. We then build and evaluate two different networks for ACE2 interacting proteins (Section C), one of which reveals connections to coagulation pathways and the other to cytokine-mediated signaling, viral response, and immunity. Finally, we discuss the potential implications of these predicted ACE2 interactions to COVID-19 pathologies and propose some specific hypotheses that emerge from this analysis.

**B. Top ERC Interactions Link ACE2 to COVID Pathologies.** To investigate protein associations of ACE2, we first determined the protein enrichment categories for its top 2% ERC proteins (based on Spearman rank correlation coefficients,  $p$ ) using the gene set enrichment package Enrichr (Xie et al., 2021) (Table 2). The top two KEGG\_2019\_Human enrichments are for complement and coagulation cascade related ( $FDR = 2.0E-03$ ) and cytokine-cytokine receptor interaction related ( $FDR = 2.0E-03$ ) terms. This finding is consistent with two hallmarks of COVID-19 pathology, abnormal systemic blood-clotting (thrombosis) and cytokine storms (Coperchini et al., 2020; Fei et al., 2020). Additionally, several terms related to viral/bacterial-specific infection are significantly enriched, such as Tuberculosis ( $FDR = 1.4E-02$ ), HPV infection ( $FDR = 1.4E-02$ ), measles ( $FDR = 2.4E-02$ ) and Hepatitis C ( $FDR = 3.1E-02$ ). Gene Ontology Biological Process also shows enrichment for tumor necrosis factor (TNF) pathways, including the signaling pathway ( $FDR = 3.9E-03$ ) and cellular responses ( $FDR = 1.6E-02$ ). Additional terms are shown in Table 2.

Enrichr Gene set	Term	FDR P-value	Odds Ratio	Gene List
KEGG_2019_Human	Complement and coagulation cascades	2.03E-03	25.9	CLU, F2RL2, F5, SERPINA5
KEGG_2019_Human	Cytokine-cytokine receptor interaction	2.03E-03	10.5	TNFSF18, IFNAR2, XCR1, IL23A, CD27, TNFSF8
KEGG_2019_Human	Tuberculosis	1.38E-02	11.0	IL23A, TRADD, MRC1, RELA
KEGG_2019_Human	Human papillomavirus infection	1.38E-02	7.6	IFNAR2, COL4A4, TRADD, COL4A5, RELA
KEGG_2019_Human	Protein digestion and absorption	1.38E-02	16.3	ACE2, COL4A4, COL4A5
KEGG_2019_Human	Pathways in cancer	1.38E-02	5.7	IFNAR2, AR, IL23A, COL4A4, COL4A5, RELA
KEGG_2019_Human	Small cell lung cancer	1.38E-02	15.8	COL4A4, COL4A5, RELA
KEGG_2019_Human	Amoebiasis	1.38E-02	15.3	COL4A4, COL4A5, RELA
KEGG_2019_Human	AGE-RAGE signaling pathway in diabetic complications	1.38E-02	14.6	COL4A4, COL4A5, RELA
KEGG_2019_Human	Toll-like receptor signaling pathway	1.39E-02	14.0	IFNAR2, TLR8, RELA
KEGG_2019_Human	Sphingolipid signaling pathway	1.85E-02	12.2	CERS4, TRADD, RELA
KEGG_2019_Human	Relaxin signaling pathway	2.18E-02	11.2	COL4A4, COL4A5, RELA
KEGG_2019_Human	Measles	2.38E-02	10.5	IFNAR2, TRADD, RELA
KEGG_2019_Human	Hepatitis C	3.05E-02	9.3	IFNAR2, TRADD, RELA
KEGG_2019_Human	Cocaine addiction	3.05E-02	19.7	PDYN, RELA
KEGG_2019_Human	PI3K-Akt signaling pathway	4.17E-02	5.5	IFNAR2, COL4A4, COL4A5, RELA
KEGG_2019_Human	Kaposi sarcoma-associated herpesvirus infection	4.17E-02	7.7	IFNAR2, TRADD, RELA
KEGG_2019_Human	Inflammatory bowel disease (IBD)	4.34E-02	14.7	IL23A, RELA
KEGG_2019_Human	Epstein-Barr virus infection	4.34E-02	7.1	IFNAR2, TRADD, RELA
KEGG_2019_Human	Adipocytokine signaling pathway	4.34E-02	13.8	TRADD, RELA
KEGG_2019_Human	RIG-I-like receptor signaling pathway	4.34E-02	13.6	TRADD, RELA
KEGG_2019_Human	Pertussis	4.85E-02	12.5	IL23A, RELA
GO_Biological_Process_2018	tumor necrosis factor-mediated signaling pathway (GO:0033209)	3.86E-03	21.0	TNFSF18, TRADD, CD27, TNFSF8, RELA
GO_Biological_Process_2018	cellular response to tumor necrosis factor (GO:0071356)	1.63E-02	13.1	TNFSF18, TRADD, CD27, TNFSF8, RELA
GO_Biological_Process_2018	immunoglobulin mediated immune response (GO:0016064)	1.63E-02	154.6	CD27, TLR8
GO_Biological_Process_2018	B cell mediated immunity (GO:0019724)	1.63E-02	154.6	CD27, TLR8
GO_Biological_Process_2018	positive regulation of NF-kappaB transcription factor activity (GO:0051092)	1.85E-02	15.6	TNFSF18, TRADD, CLU, RELA
GO_Biological_Process_2018	I-kappaB kinase/NF-kappaB signaling (GO:0007249)	2.26E-02	26.3	TRADD, TLR8, RELA
GO_Biological_Process_2018	cytokine-mediated signaling pathway (GO:0019221)	3.11E-02	5.7	TNFSF18, IFNAR2, IL23A, TRADD, CD27, TNFSF8, RELA
GO_Biological_Process_2018	regulation of inflammatory response (GO:0050727)	3.11E-02	11.9	ACE2, IL23A, PLA2G7, RELA
GO_Biological_Process_2018	positive regulation of defense response (GO:0031349)	3.26E-02	20.0	IL23A, TLR8, PLA2G7
WikiPathways_2019_Human	Complement and Coagulation Cascades WP558	2.38E-02	25.8	CLU, F5, SERPINA5
WikiPathways_2019_Human	EBV LMP1 signaling WP262	4.24E-02	44.2	TRADD, RELA
WikiPathways_2019_Human	Toll-like Receptor Signaling Pathway WP75	4.24E-02	14.2	IFNAR2, TLR8, RELA
WikiPathways_2019_Human	Toll-like Receptor Signaling WP3858	4.36E-02	32.0	TLR8, RELA
WikiPathways_2019_Human	miRNAs involvement in the immune response in sepsis WP4329	4.95E-02	26.5	TLR8, RELA
WikiPathways_2019_Human	Regulation of toll-like receptor signaling pathway WP1449	4.96E-02	10.4	IFNAR2, TLR8, RELA

**Table 2: Enrichment categories for ACE2's top 2% proteins by ERC.** Key enrichments include complement and coagulation cascades, cytokine-cytokine signaling, and different pathogen infections.

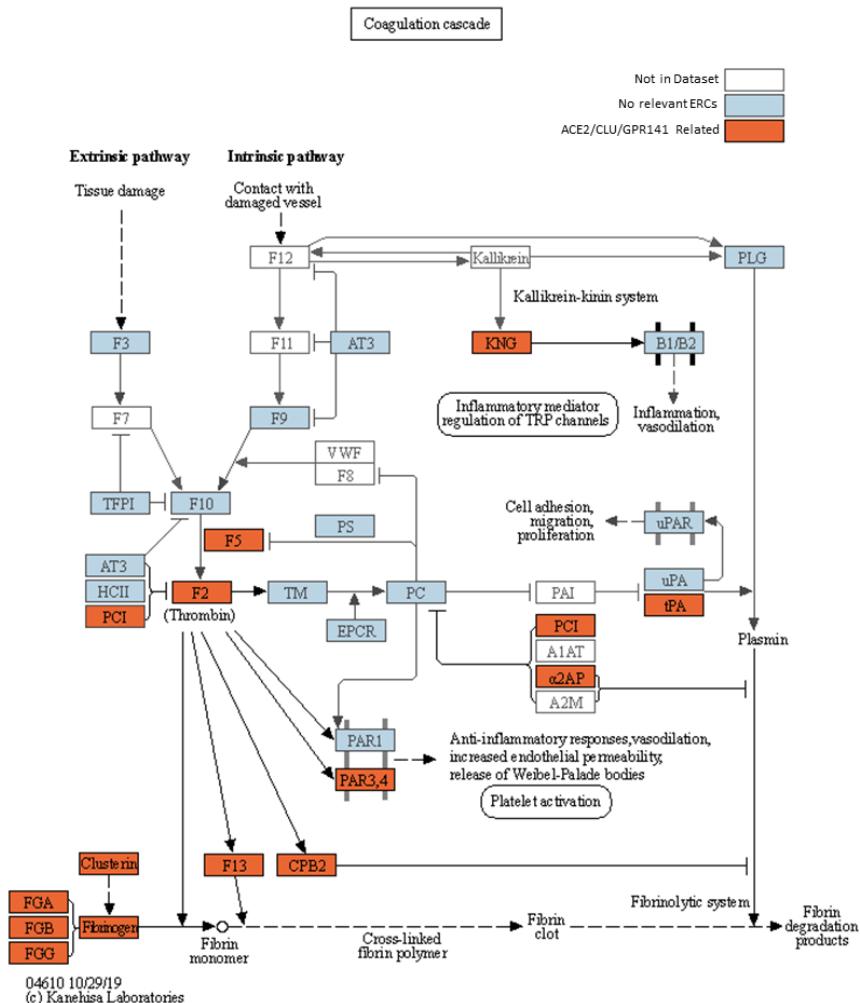
The ACE2 ERC analysis indicates that ACE2 is “coevolving” with proteins involved in the complement and coagulation pathways, cytokine signaling, TNF, and pathogen response pathways. Here, we summarize results and background information on some of the key proteins

among ACE2's ERCs (more extended summaries of each protein are in the Supplementary Text).

Among ACE2's strongest ERCs are proteins involved in immunity. For example, XCR1 (X-C Motif Chemokine Receptor 1) is ACE2's 2nd top-ranked ERC ( $\rho = 0.67$ , FDR = 6.2E-05). It is a chemokine XCL1 receptor involved in immune response to infection and inflammation (Lei & Takahama, 2012). Strikingly, the Severe Covid-19 GWAS Group (2020) detected a small genomic region containing six genes that significantly associate with severe COVID-19, one of which is XCR1. Our finding that XCR1 is ACE2's 2nd highest ERC interactor lends independent support for a relationship between COVID-19 and XCR1. Furthermore, it suggests that an interaction between ACE2 and XCR1 could be involved in COVID-19 pathologies. To our knowledge, there are no other reports of interactions between these two proteins.

Another striking connection of ACE2 ERC to immunity is through IFNAR2 (Interferon alpha/beta receptor 2), which has a highly significant ACE2 ERC correlation ( $\rho = 0.62$ , FDR = 6.1E-04). IFNAR2 forms part of an important receptor complex with IFNAR1 (Thomas et al., 2011) involved in interferon signaling through the JAK/STAT pathway to modulate immune responses. IFNAR2 has been implicated in severe COVID-19, based on mendelian randomization, genome-wide associations, and gene expression changes (Liu et al., 2021; Pairo-Castineira et al., 2021). Our data provide independent support for a role, possibly mediated through ACE2 interactions. Interferon pathways are important in antiviral defense, but also can contribute to cytokine storms and COVID-19 pathologies (McKechnie & Blish, 2020). Other immune-related proteins with high ERC connections to ACE2 include TLR8 (Toll-like Receptor 8), FAM3D (FAM3 metabolism regulating signaling molecule D), and PLA2G7 (phospholipase A2 group VII).

Coagulation pathway proteins figure prominently in ACE2 ERC-predicted protein interactions (Table 3, Fig. 2). This is reflected both in significant enrichment for coagulation cascade proteins in the top 2% strongest ACE2 ERCs (Table 2) and the strong reciprocal rank network for ACE2 (Section C, Fig. 3). The finding has obvious potential implications to a hallmark pathology of COVID-19, systemic coagulopathy (Wright et al., 2020; Medcalf, Keragala & Myles, 2020). A list of coagulation and blood-related proteins associated with ACE2 is presented in Table 3. Among ACE2's top 2% ERCs associated with coagulation pathway are Coagulation Factor V (F5), Protein C inhibitor (SERPINA5 aka PCI), and Thrombin Receptor 2 (F2RL2) (Table 1).



**Figure 2: KEGG Coagulation Pathway.** KEGG Coagulation cascade pathway (Kanehisa & Goto, 2000), with ACE2-CLU-GPR141 associated proteins (based on presence on any of their top 2% ERCs or in the ACE2 CRR network) indicated in orange. The KEGG pathway has been supplemented to indicate the three fibrinogen proteins and clusterin associations previously discussed. Note the alternate protein names: PAR3,4 = F2RL2 & F2RL3 = Thrombin receptors; α2AP = Alpha-2-antiplasmin = SERPINF2; PLAT = tPA, and PCI = SERPINA5 = Protein C Inhibitor.

Also relevant to coagulopathy are Clusterin (CLU) and the orphan G protein-coupled receptor 141 (GPR141). The chaperone protein CLU has a soluble form that circulates in the blood and is part of the “cleaning squad” that clears misfolded extracellular proteins for delivery to lysosomes and degradation (Itakura et al., 2020; Sánchez-Martín & Komatsu, 2020). It is the 3rd highest ACE2 ERC ( $p = 0.63$ , FDR = 1.5E-04), and these two proteins show strong reciprocal ranks (3, 8), likely supporting biological interactions. Relevant to this point is that both ACE2 and CLU have soluble forms that circulate in the blood (Itakura et al., 2020). Of direct relevance to COVID-19 and possible ACE2-CLU protein interactions, Singh et al. (2021) found in cells infected with different coronaviruses (SARS-CoV-2, SARS-CoV, and MERS-CoV), only two genes were found to be differentially expressed in all three, with CLU being one.

CLU's top 2% strongest ERCs show highly significant enrichment for terms relating to coagulation cascades and clot formation (Supplementary File 3, e.g. "Complement and coagulation cascades", FDR = 6.3E-12), as well as significant terms that are relevant to immunity, such as "Immune system" (FDR = 4.8E-03) and "activated immune cell type" (FDR = 3.4E-05). Among its top ERC proteins relevant to coagulation process are Coagulation Factor V (F5,  $\rho$  = 0.67, FDR = 9.1E-06, rank 3), Fibrinogen Gamma chain (FGG,  $\rho$  = 0.59, FDR = 1.7E-04, rank 18), Coagulation Factor XIII B chain (F13B,  $\rho$  = 0.63, FDR = 2.8E-05, rank 19), and Fibrinogen Alpha chain (FGA,  $\rho$  = 0.57, FDR = 2.9E-04, rank 27) (Fig. 3, Supplementary File 1). Notably, fibrinogen is a major binding "client" of Clusterin in stressed plasma (Wyatt & Wilson, 2010). Little is known about GPR141; however, the ERC analysis suggests an important role in blood coagulation. Among GPR141's top ERC proteins relevant to coagulation process are Kininogen 1 (KNG1,  $\rho$  = 0.60, FDR = 9.3E-04, rank 5), Plasminogen Activator (PLAT,  $\rho$  = 0.58, FDR = 6.5E-04, rank 6), Thrombin (Coagulation Factor II or F2,  $\rho$  = 0.58, FDR = 6.5E-04, rank 7), Fibrinogen Beta chain (FGB,  $\rho$  = 0.57, FDR = 6.5E-04, RR 11, 11), Complement C1s (C1S,  $\rho$  = 0.54, FDR = 1.6E-03, rank 22), F2R-like thrombin (also called trypsin receptor 3; F2RL3,  $\rho$  = 0.52, FDR = 2.6E-03, rank 37), and Coagulation Factor V (F5,  $\rho$  = 0.52, FDR = 1.7E-03, rank 39) (Fig. 2, Supplementary File 1).

GPR141 has a highly significant ERC to CLU, with these two proteins being each other's first ranking ERCs ( $\rho$  = 0.68, FDR = 9.1E-06, RR 1,1). The pattern suggests a strong biological interaction, although none is described in the literature. The result supports investigating functional interactions between CLU and GPR141, based upon their high ERC and reciprocal ranks. Our network analysis (Section C) further supports extensive interconnections among ACE2, Clusterin, GPR141, and coagulation pathway proteins, implicating the protein interaction pathway as a possibly significant contributor to disruption of coagulation in COVID-19 disease. Coagulation cascade proteins found in the ACE2's top 2% ERCs, ACE2 reciprocal rank network, and Clusterin-GPR141 associated proteins are highlighted in Figure 2.

Androgen Receptor (AR,  $\rho$  = 0.57, FDR = 8.8E-04, rank 13) is the receptor for the male hormone androgen. It plays a major role in reproductive system development, somatic differentiation, and behavior (Matsumoto et al., 2008). Androgen-AR signaling induces ACE2 (Wu et al., 2020), while knockdowns of AR result in downregulation of ACE2 (Samuel et al., 2020). AR agonists also reduce SARS-CoV-2 spike protein-mediated cellular entry (Deng et al., 2021). Additionally, AR is associated with COVID-19 comorbidities (Dolan et al., 2020), and recently implicated in the severity of COVID-19 in women with polycystic ovarian syndrome, a disorder associated with high androgen levels and androgen sensitivity (Gotluru et al., 2021). Our ERC finding indicating ACE2 and AR coevolution suggests regulatory feedback between these two proteins, which could be relevant to COVID-19 severity and other sex differential pathologies, such as cardiovascular disease (Viveiros et al., 2021).

Other notable significant ACE2 ERCs (Table 1) include Metabolism regulating signaling molecule D (FAM3D), Transmembrane-protein 63C (TMEM63C); Collagen Type IV Alpha 4 (COL4A4), L1 cell adhesion molecule (L1CAM), and ITPRIP-like 2 (ITPRIPL2). More detailed information on these and other proteins mentioned in this section is provided in Section C and the Supplementary Text.

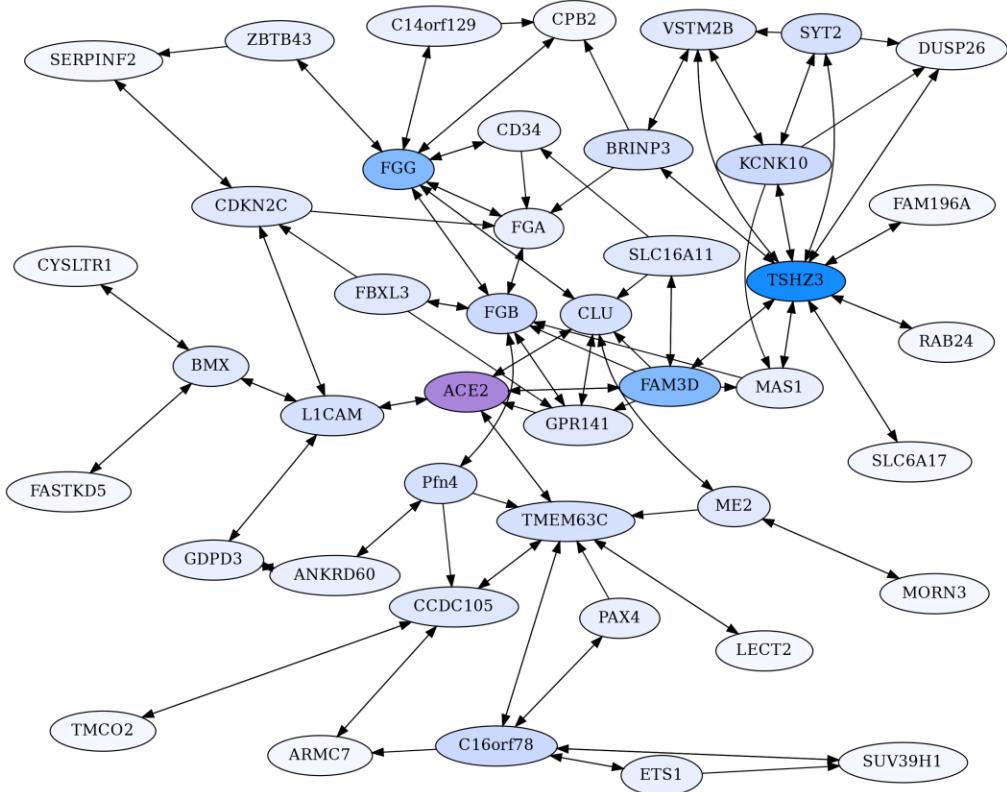
**C. ERC Reciprocal Rank Networks Implicate Coagulation Pathways and Immunity.** As mentioned previously, two proteins with a significant evolutionary rate correlation (ERC) may often “rank” each other differently in their respective top ERC connections. This occurs because some proteins have more extensive ERC connections than others. High reciprocal ERC ranks between protein pairs may be more indicative that they are under strong coevolutionary pressure in their sequence and function. We have thus found it useful to evaluate these reciprocal rank connections as a network. The rationale is that such proteins are likely to be reciprocally evolving (“coevolving”). To build reciprocal rank networks, we use protein pairs that reciprocally share ranks less than or equal to 20 (RR20), which are the top one percent for each protein’s respective ERC set.

A core ACE2 reciprocal rank network was generated by building reciprocal rank connections (RR20) outward of ACE2, to provide a backbone set of RR20 protein connections. The backbone was expanded on by adding the RR20 connections of the non-ACE2 backbone proteins. Unidirectional ERCs ( $\leq$  rank 20) were then added between proteins within the RR set to produce an ACE2 Core Reciprocal Rank (CRR) Network (Fig. 3). The network is designed to capture features of ACE2’s protein interactions as revealed by the strong reciprocal evolutionary correlations among proteins.

ACE2 also has highly significant ERCs to proteins that do not rank ACE2 within their top 1% of ERCs, due to those proteins having more protein interactions with higher ERCs. A second network was therefore generated using ACE2’s top ten unidirectional ERCs, followed by calculating the RR20 associations for those proteins. This second network is referred to as the ACE2 Unidirectional Reciprocal Rank (URR) Network (Fig. 4).

These are presented below. In general, the reciprocal ranks analysis lends credence to our proposition that ERCs reveal real biological interactions, as well as providing predictions for novel protein interactions possibly of importance to COVID-19 pathologies and protein-interaction networks.

**C1. The ACE2 Core Reciprocal Rank (CRR) Network.** The CRR network (Fig. 3) is designed to capture essential features of ACE2’s protein interactions as revealed by the strong reciprocal correlations among proteins.



**Figure 3: ACE2 Centric Reciprocal Rank (CRR) Network.** Proteins with ERC reciprocal ranks  $\leq 20$  are shown by double-headed arrows, and unidirectional ranks  $\leq 20$  connecting to the RR backbone are indicated by single-headed arrows. ACE2 has extensive connections to coagulation proteins mediated primarily through Clusterin (CLU) and GPR141. ACE2 is highlighted in purple, and blue shading intensity indicates the level of reciprocal connectivity for different proteins.

The most striking aspects of the ACE2 CRR Network are extensive connections to the coagulation pathway and blood-associated proteins (Fig. 3, Table 3). This could be relevant to COVID-19 due to extensive clotting pathologies and stroke associated with COVID-19 (Bonaventura et al., 2021), as well as microvascular clotting and the apparent shut-down of fibrinolysis (Wright et al., 2020). Extensive blood coagulation of COVID-19 patients can even lead to clogging of dialysis equipment (Rabb, 2020). This hallmark pathology of COVID-19 indicates a disruption in coagulation and fibrinolysis pathways, and our findings of extensive network connections between ACE2 and coagulation-fibrinolysis pathway proteins could be relevant. The predicted novel protein interactions detected here may also have implications more generally to circulatory system homeostasis, including regulation of blood pressure and coagulation.

Name	Full Name	Brief Description
ACE2	Angiotensin-Converting Enzyme 2	Catalyzes the cleavage of angiotensin I to angiotensin 1-9 and angiotensin II to angiotensin 1-7 (Burrell et al., 2004)
FGA	Fibrinogen alpha chain	Bind to FGB and FGG to form fibrinogen, used to form blood clots (Mosesson, 2005)
FGB	Fibrinogen beta chain	Bind to FGA and FGG to form fibrinogen, used to form blood clots (Mosesson, 2005)
FGG	Fibrinogen gamma chain	Bind to FGA and FGB to form fibrinogen, used to form blood clots (Mosesson, 2005)
CPB2	Carboxypeptidase B2	Inhibits fibrinolysis (Leenaerts et al., 2018)
SERPINF2	Serpin family F member 2 (alpha-2-antiplasmin)	Inhibits Plasmin, a protein involved in fibrinolysis (Kanehisa & Goto, 2000)
CD34	CD34 molecule	Associated with hematopoiesis and stem cells (Fina et al., 1990)
CLU	Clusterin	Binds to Fibrinogen (Wyatt & Wilson, 2010)
MAS1	MAS1 Proto-Oncogene, G Protein-Coupled Receptor	Receptor for angiotensin-(1-7) (Burrell et al., 2004)
FAM3D	FAM3 Metabolism Regulating Signaling Molecule D	Implicated in inflammatory responses in the gastrointestinal tract and is a chemoattractant for neutrophiles and monocytes (Peng et al., 2016)
GPR141	G Protein-Coupled Receptor 141	High expression in blood, granulocytes, Kupfer cells, and macrophages (Stelzer et al., 2016)
TMEM63C	Transmembrane Protein 63C	Interacts with angiotensin II (Eisenreich et al., 2020)
LECT2	Leukocyte Cell-derived Chemotaxin 2	Involved in macrophage activation, insulin resistance and diabetes, and neutrophil chemotaxis (Yamagoe et al., 1996; Zhang et al., 2018; Takata et al., 2021)
ETS1	ETS proto-oncogene 1, transcription factor	Transcription factor involved in cytokine/chemokine processes and angiogenesis (Stelzer et al., 2016)
ZBTB43	Zinc Finger and BTB Domain containing 43	Associated with Diamond-Blackfan Anemia 4, in which the bone marrow is unable to make enough red blood cells to carry oxygen (Stelzer et al., 2016)
COL4A4	Collagen Type IV Alpha 4	Subunit of Collagen Type 4, which are a part of the basement membrane which resides between epithelial cells (Stelzer et al., 2016)
F13B	Coagulation Factor XIII B chain	Stabilizes F13A subunits, while it does not have enzymatic abilities it is thought to be a plasma carrier molecule (Stelzer et al., 2016)
AMOT	Angiomotin	Associated with angiogenesis and endothelial cell movement (Bratt et al., 2005; Aase et al., 2007)
PDYN	Prodynorphin	Inhibits vasopressin secretion (Yamada et al., 1988)

**Table 3: ACE2-derived Coagulation and Blood-related Proteins.** Coagulation and blood-related proteins in the ACE2 CRR and URR Networks as well as the top 1% ACE2 ERC list.

ACE2 connects to coagulation pathway proteins through F5, CLU, FAM3D, and GPR141 (Fig. 2, Fig. 3). CLU-GPR141 form a high RR ERC (ranks 1,1), strongly suggesting coevolution of these proteins and physical/functional interactions. Both CLU and GPR141 then connect to the fibrinogen proteins FGB and FGG. FGA, FGB, and FGG are the three protein components that make up fibrinogen, which during the clotting process are converted into fibrin monomers, which subsequently cross-link to form the fibrin clot (Mosesson, 2005). All three proteins form an RR20 triad, indicating protein coevolution. FGG is a hub for RR ERCs to several other proteins (e.g. CD34, CPB2, C14or129, and ZBTB43). ZBTB43 is noteworthy, as it is associated with the blood diseases Diamond-Blackfan Anemia 4 and Hemochromatosis Type 2 (Stelzer et

al., 2016). The former disrupts red blood cell formation in the bone marrow and the latter causes iron accumulation in the body. In terms of tissue distribution, ZBTB43 is enhanced in bone marrow (Uhlén et al., 2015). Cellularly, it is found mainly in nucleoplasm and nucleoli, suggesting regulatory functions, as might be expected for a transcription factor-like zinc finger domain protein. Most noteworthy, Mamoor (2020) has shown that ZBTB43 is differentially expressed in human microvascular endothelial cells and human cell cultures infected with coronaviruses (e.g MERS-CoV and human coronavirus 229E). So, this is yet another member of the ACE2 protein Network which is implicated in coronavirus infection. In turn, ZBTB43 has a RR connection with SERPINF2, which enhances clotting by inhibiting plasmin, an enzyme that degrades fibrin, the main component of clots. Mutations in SERPINF2 can cause severe bleeding disorders and upregulation of SERPINF2 is implicated in COVID-19 patient thrombosis (Jain et al., 2021; Lazzaroni et al., 2021). In turn, CPB2 (Carboxypeptidase B2) is a thrombin-activated inhibitor of fibrinolysis, and therefore enhances clotting stability (Leenaerts et al., 2018), and also plays a role in activating the complement cascade (Morser et al., 2018; Leung & Morser, 2018).

FAM3D is a cytokine for neutrophils and monocytes in peripheral blood which may interact with ACE2 based on their reciprocal ranking. ACE2 is its 2nd ranking ERC. Although ACE2 does not have a significant ERC to F13B (also known as Coagulation Factor XIII B Chain), it is FAM3D's top-ranking ERC. F13B functions to stabilize clotting through cross-linking of fibrin (Stelzer et al., 2016). Thus, the predicted interaction of FAM3D and F13B may be relevant to the coagulation pathway.

Blood pressure and vasoconstriction regulation also show functional enrichment in the CRR network. Naturally, ACE2 is a crucial component of the Renin-Angiotensin System (RAS), which converts angiotensin II to angiotensin (1-7). This, in turn, binds to the MAS1 receptor, promoting vasodilation and reduced blood pressure. As seen in Figure 2, MAS1 is part of the ACE2 CRR network. Although not significantly correlated with ACE2 directly, it has significant RR connection to TSHZ3 ( $\rho = 0.52$ , FDR = 7.8E-03, ranks 11, 4) and is FAM3D's 19th ranking ERC ( $\rho = 0.49$ , FDR = 1.5E-02). Biologically MAS1 and ACE2 are key elements promoting vasodilation in the renin-angiotensin system (RAS) (Burrell et al., 2004). Thus, the ERC RR network detects biologically significant connections of ACE2 to RAS signaling via the MAS1 receptor of angiotensin-(1-7). Samavati & Uhal (2020) posit that the loss of ACE2 due to SARS-CoV-2 infection reduces MAS1 signaling and increases AT1 & AT2 signaling via higher levels of angiotensin 2, promoting vasoconstriction, fibrosis, coagulation, vascular and cardio injury, and ROS production. Similar arguments are made by Sriram & Insel (2020). ACE2 and MAS1 do not have a signature of protein coevolution, even though they interact indirectly biologically through the short seven amino acid signaling peptide Ang (1-7). In contrast, MAS1 has a significant RR with TSHZ3 (mentioned above). A biological connection between these proteins is not obvious, although the high ERC reciprocal ranks suggest possible interactions worth further investigation. Additionally, TMEM63C is one of four proteins that form a reciprocal rank

ERC association with ACE2 (Figure 2). It functions in osmolarity regulation and like ACE2, interacts with angiotensin II, possibly reducing damage to kidney podocytes (Eisenreich et al., 2020).

FBXL3 has a RR20 connection to FGB and ranks GPR141 in its top 2%. This protein is a component of circadian rhythm regulation (Busino et al., 2007). Many aspects of the cardiovascular system have circadian cycling such as heart rate, blood pressure, and fibrinolysis (Reilly, Westgate & FitzGerald, 2007). Endogenous oscillators in the heart, endothelial cells, and smooth muscles may play significant roles in these cycles (Reilly, Westgate & FitzGerald, 2007), and the CRR network suggests that interactions between FBXL3 and FGB could play a role in circadian aspects of fibrinolysis.

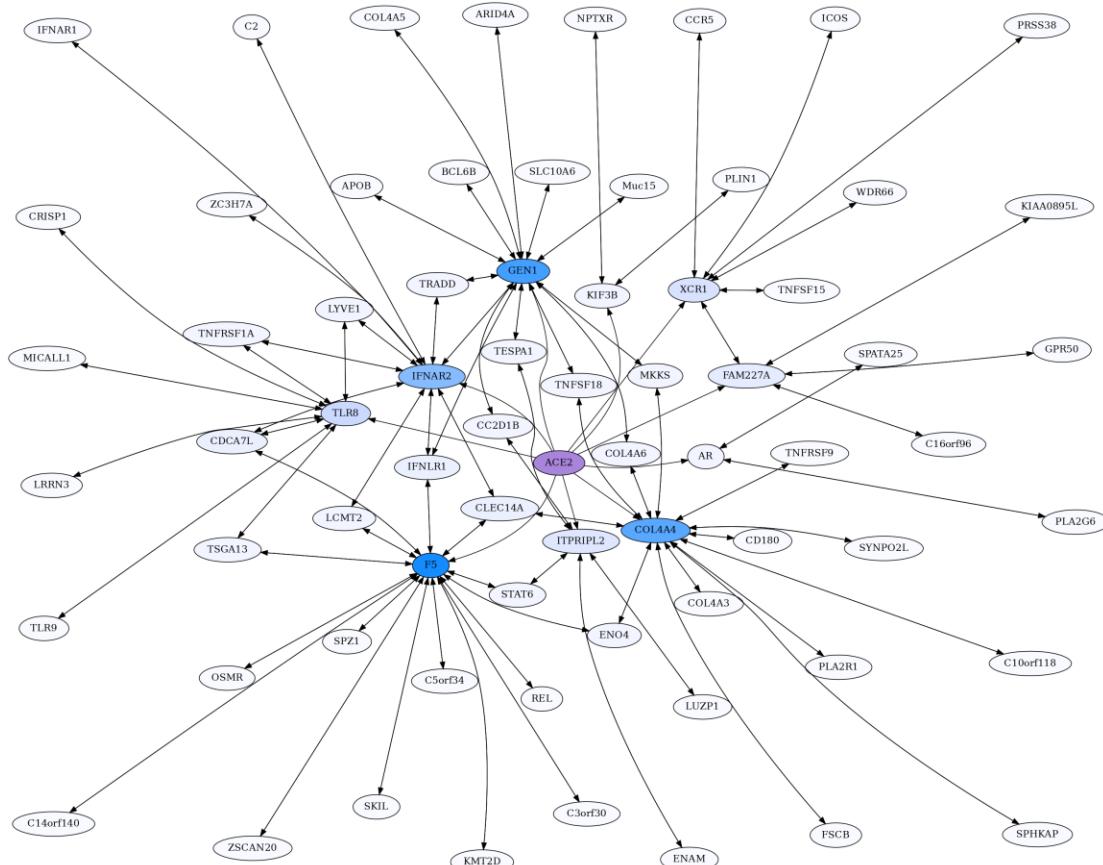
CD34 (Hematopoietic Progenitor Cell Antigen CD34) is believed to be an adhesion protein for hematopoietic stem cells in bone marrow and for endothelial cells (Fina et al., 1990). Our ERC analysis indicates connections to coagulation pathway proteins and lipoproteins. In addition to its RR association with FGG ( $\rho = 0.60$ , FDR = 2.2E-04, ranks 18,9), CD34 also forms significant reciprocal rank correlations with coagulation factor F2 ( $\rho = 0.69$ , FDR = 7.9E-06, ranks 1,6), lipoprotein APOE ( $\rho = 0.64$ , FDR = 6.0E-05), lipid droplet-associated protein PLIN1 ( $\rho = 0.64$ , FDR = 1.1E-04, ranks 8,7), and inflammation associated pentraxin protein PTX3 ( $\rho = 0.65$ , FDR = 6.8E-05, ranks 3,11) (Supplementary File 1). As expected from these protein associations, CD34's top enriched term is to complement and coagulation cascade (FDR = 1.4E-08). There is also enrichment for HUVEC cells (FDR = 3.1E-05) and Blood Plasma (FDR = 1.7E-04) (Supplementary File 3).

Additional proteins of interest are discussed further in the supplementary materials, including TSHZ3 (a key regulator of airflow and respiratory rhythm control) and L1CAM (involved in nervous system development and vascular endothelial cell differentiation from stem cells).

Consistent with the descriptions above, the CRR network shows enrichment (full enrichment table in Supplementary File 3) for negative regulation of blood coagulation (FDR = 4.3E-08), platelet alpha granule-related terms (FDR = 1.7E-05), plasma cell (FDR = 8.3E-4) and blood clot (FDR = 4.5E-02). These enrichments indicate that the network involves protein interactions related to blood clotting pathways. There are also several significantly enriched terms which are driven in part by ACE2, such as regulation of systemic arterial blood pressure by renin-angiotensin (FDR = 1.6E-03), metabolism of angiotensinogen to angiotensin (FDR = 6.9E-03), regulation of blood vessel diameter (FDR = 1.5E-02), and renin-angiotensin system (FDR = 1.8E-02).

**C2. The ACE2 Unidirectional Reciprocal Rank (URR) Network.** ACE2 also has highly significant ERCs with interacting proteins that are unidirectional, meaning that ACE2 ranks these proteins in its top 2%, but the partner protein does not rank ACE2 within its top 2% due to higher ERC correlations with other partners (Table 1). Some of ACE2's highest-ranking proteins

fall into this category, including GEN1 (rank 1), XCR1 (2), IFNAR2 (5) KIF3B (6), and ITPRIPL2 (7), FAM227A (8), TLR8 (9), COL4A4 (10), F5 (12), and AR (13). To focus on strong protein connections in this set, we took the top ten proteins with unidirectional ERCs for ACE2 and then added their reciprocal rank 20 (RR20) partners. The resulting ACE2 Unidirectional Reciprocal Rank (URR) Network contains 69 proteins (Fig. 4).



**Figure 4: ACE2 Unidirectional Reciprocal Rank (URR) Network.** ACE2's top 10 unidirectional ERC proteins for a web of reciprocal rank (RR20) connections. The network is particularly enriched for cytokine signaling and immunity. Highly interconnected proteins include COL4A5, F5, GEN1, and IFNAR2. ACE2 is highlighted in purple, and blue shading intensity indicates the level of reciprocal connectivity for different proteins.

Notable in the network are many proteins involved in immunity and cytokine signaling, such as IFNAR2 (Interferon alpha/beta receptor 2), XCR1 (X-C Motif Chemokine Receptor 1), and ICOS (Inducible T Cell Costimulator). There are also Toll-Like Receptors TLR8 and TLR9, which stimulate innate immune activity (Forsbach et al., 2011), and Tumor Necrosis Factor related proteins such as TNFSF18, TNFSF15, TNFRSF9, and TNFRSF1A.

Enrichment analysis of the URR network generates 72 significant terms (Supplementary File 3). The network is highly enriched for cytokine-cytokine receptor interaction (FDR = 6.5E-06), I-kappaB kinase/NF-kappaB signaling (FDR = 1.6E-06), necroptosis (FDR = 3.3E-03), viral infections, such as Human Papillomavirus (FDR = 5.7E-04) and Herpes virus (FDR = 3.5E-03),

JAK-STAT and PI3K-AKT signaling pathways, Toll-like receptor signaling, and immune system Homo sapiens (FDR = 3.7E-03).

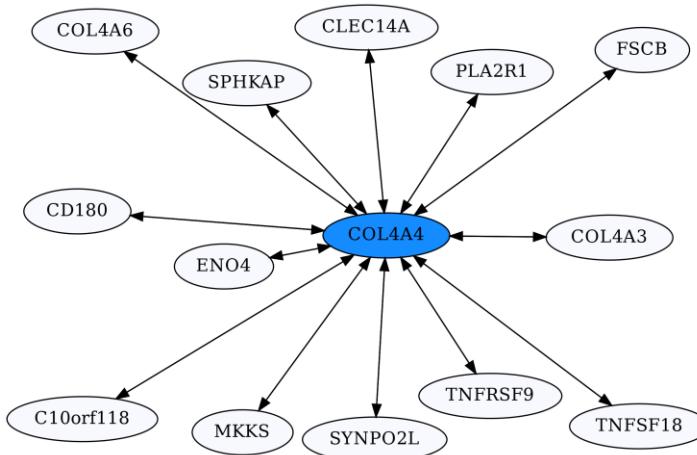
XCR1 is the 2nd highest ACE2 ERC. It is the receptor for chemokine XCL1, which is produced in response to infection and inflammation, and during development of regulatory T cells (Lei & Takahama, 2012). Furthermore, XCR1 maps to a region implicated in severe COVID-19 by a genome-wide association study (Severe Covid-19 GWAS Group, 2020). As seen in Figure 4, XCR1 forms a RR subnetwork with six other proteins (ICOS, CCR5, WDR66, TNFSF15, PRSS38, and FAM227A), three of which are known to be involved in immunity. ICOS (Inducible T Cell Costimulator) is reciprocally evolving with XCR1 based on their ERC interaction. It is an inducible T Cell stimulator that is essential for T helper cell responses (Hutloff et al., 1999; Tafuri et al., 2001). In addition, ICOS signaling is impaired in COVID-19 patients requiring hospitalization (Hanson et al., 2020). The high ERC between ACE2 and XCR1 and high reciprocal ranks of XCR1 to ICOS suggests that the disruption of an ACE2-XCR1 interaction could have a contributory role in COVID-19. C-C Motif Chemokine Receptor 5 (CCR5) forms a significant RR ERC with XCR1 as well. Several studies have implicated CCR5 variation and expression to be associated with COVID-19 severity (Gómez et al., 2020; Hubacek et al., 2021; Kasela et al., 2021), while others have not (Bernas et al., 2021). TNFSF15 is a third immune response protein in the XCR1 RR subnetwork that shows elevated expression in patients with severe COVID-19 (Jain et al., 2021). We recognize that the involvement of these immune-related proteins in COVID-19 does not require an effect mediated through ACE2. Instead, their protein evolutionary correlations suggest that ACE2 may play a contributory role to COVID-19, possibly through XCR1-related pathways, as suggested by the network analysis.

IFNAR2 is another protein that is highly correlated with ACE2 ( $\rho = 0.62$ , FDR = 6.1E-04) and is also implicated in severe COVID-19 by GWAS and expression data (Liu et al., 2021; Pairo-Castineira et al., 2021). It has RR20 ERCs with ten other proteins and is embedded in a complex web of interactions with members of the ACE2 network. Here we draw attention to a few key features. Notably, IFNAR2 and IFNAR1 are RR partners, as expected given that they combine to form the IFN-alpha/beta receptor, which is the receptor for both alpha and beta interferons. IFNAR2 forms a high RR relationship with TNFRSF1A ( $\rho = 0.84$ , FDR = 4.8E-12, 1,1 reciprocal ranks). This protein is the receptor for TNF $\alpha$  and the pathway affects apoptosis and inflammation regulation. Jin et al. (2015) found that ACE2 deletion increases inflammation through TNFRSF1A signaling, lending further support to a functional association between ACE2 and this protein.

GEN1 is the highest-ranking ACE2 ERC protein ( $\rho = 0.67$ , FDR = 4.2E-05), and it functions as a resolvase of Holliday junctions and a DNA damage checkpoint signaling (Chan & West, 2015). Frankly, we are perplexed by the functional significance of ACE2-GEN1 correlated evolution. As observed in the ACE2 network, GEN1 is a highly interconnected protein, with 14 RR20 connections in the network. This result suggests that GEN1 may have additional functions

beyond DNA replication. Indeed, although its second-highest RR is to CC2D1B (2,1), a protein involved in mitosis, its highest RR is to Interferon Lambda Receptor 1 (IFNLR1), with an impressive Spearman correlation of  $\rho = 0.89$  (FDR = 6.2E-17). As IFNLR1 binds cytokine ligands and stimulates antiviral response, this suggests some feedback mechanism between GEN1 and the immune system, possibly related to its functional role in DNA damage checkpoint signaling. Indeed, its top 2% ERCs show enrichment for multiple viral infection terms (Supplementary File 3). Therefore, it appears that GEN1 has a “hidden life” that ERC analysis suggests warrants exploration.

The Collagen Type IV A4 subnetwork (Fig. 4, Fig. 5) lends further credence to the view that ERCs can detect proteins with likely binding partners. COL4A4 is a component of the Collagen Type IV protein complexes in basement membranes in the extracellular matrix of various tissues, including the kidney glomerulus and vascular endothelial cells, and lung alveoli (Myllyharju & Kivirikko, 2001). COL4A4, COL4A3, and COL4A5 complex with each other in the basement membranes of kidney glomeruli – mutations in these COL4A proteins are known to cause different kidney disorders (Torra et al., 2004; Wiradjaja, DiTommaso & Smyth, 2010). Consistent with their expected binding, COL4A4 and COL4A3 are each other’s reciprocal best partners (ranks 1,1) and highly correlated with each other ( $\rho = 0.88$ , FDR = 4.4E-16). Both show highly significant ERCs to COL4A6 (rank 6,5 for COL4A4  $\rho = 0.83$ , FDR = 2.1E-12; rank 22,30 for COL4A3  $\rho = 0.78$ , FDR = 1.6E-10). Thus, evolutionary rate correlations show highly significant ERCs among Collagen Type IV proteins known to physically interact. A future direction is to use ERCs to more precisely define predicted coevolving protein segments, which could be used to inform docking simulations and experimental studies.

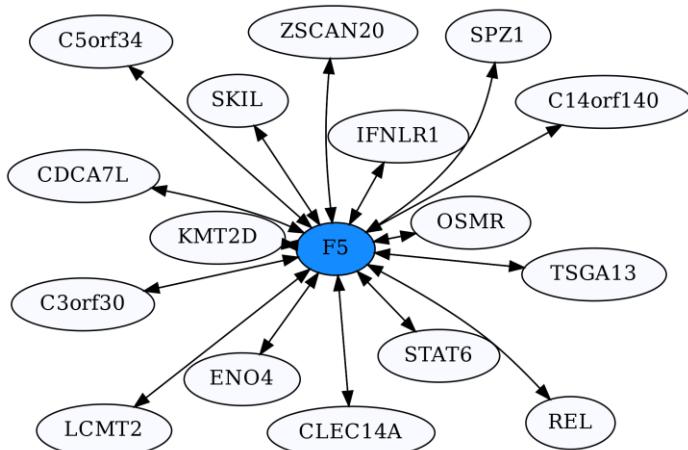


**Figure 5: COL4A4-Centric RR20 Network.** This network detects reciprocal ERCs of different proteins to COL4A4, including other COL4A proteins known to form complexes with COL4A4.

COL4A5 also has significant ERCs to COL4A3 ( $\rho = 0.71$ , FDR = 2.2E-08) and COL4A4 ( $\rho = 0.71$ , FDR = 1.7E-08), but these do not qualify as RR20 due to the large number of high ERCs for COL4A5. Interestingly, COL4A5-MUC15 are top-ranking partners (ranks 1,1) with a very

high ERC ( $\rho = 0.89$ , FDR = 3.2E-16). MUC15 is a cell surface protein that is believed to promote cell-extracellular matrix adhesion and it is implicated in affecting influenza infection (Chen et al., 2019), which may increase its relevance in the context of COVID-19 infection. ERCs may help to inform candidate domains within each protein that are involved in their expected binding affinity.

Coagulation Factor V (F5) is known for its role in the coagulation cascade. However, F5 is a highly ERC-connected protein, with 43 proteins ranking it in their respective top 5 highest ERCs. This connectedness is also reflected in the RR20 network shown below (Fig. 6). F5 has 16 RR20 connections out of the maximum 20 possible. Although F5 is a vital protein in the coagulation cascade, its top 16 RR connections indicate immune functions, including Interferon λ receptor 1 (IFNLR1; RR 4,10) and Oncostatin M Receptor (OSMR; RR 1,4). This is reflected in the enrichments among its 16 RR proteins for the JAK-STAT signaling pathway (FDR = 8.7E-03) and response to cytokine (FDR = 2.5E-02). Similarly, the F5 top 2% ERC show enrichments for 54 terms (Supplementary File 3); notably many related to inflammatory response (FDR = 1.1E-03) and the complement system (FDR = 8.4E-03). The functions of several of F5's RR20 partners are not well known, such as C14orf140 and C5orf34. Their top 2% enrichment suggests cytokine receptor activity (FDR = 2.7E-02) for C14orf140, and Human Complement System (FDR = 1.9E-03) and cytokine receptor activity (FDR = 2.1E-02) for C5orf34. In conclusion, F5 appears to have a “secret life” of strong protein interactions reflecting moonlighting functions with extensive signaling or modulation roles beyond coagulation regulation.



**Figure 6: Coagulation Factor V-centric RR20 Network.** The network captures strong reciprocal ERCs between F5 and proteins related to immune function such as IFNLR1.

**D. ERCs and Protein Interactions.** We postulate that ERCs detect proteins that are coevolving due to functional interactions. Furthermore, we propose that physical binding is an important mechanism contributing to significant ERCs between proteins. This is consistent with anecdotal observations from this study of high reciprocal rank ERCs among the fibrinogen components

FGA, FGB, & FGG, the Collagen Type IVA proteins COL4A4, COL4A3, and COL4A6 proteins, and Interferon alpha/beta proteins IFNAR2 and IFNAR1.

To further investigate the role of binding affinity, we examined the mammalian protein complex database CORUM (Giurgiu et al., 2019) to determine whether significantly higher Spearman rank correlations ( $\rho$  values) are found among proteins within known protein complexes. A set of 139 protein complexes (excluding those with overlapping proteins) were identified which contain at least two members from our ERC data set, for a total of 258 pairwise comparisons. We compared the  $\rho$  values of within complex proteins to the median values for proteins outside the complex and found that Spearman rank correlations of within complex proteins were significantly higher than its between complex values according to Wilcoxon matched signs rank tests (WMRST) under a significance level of  $\alpha = 0.05$  ( $p = 5.2\text{E-}04$ ), with a median increase of 6.3% (Supplementary File 11). Many of the complexes contain large numbers of proteins, reducing the probability of direct physical contact between individual members. We therefore also analyzed only proteins from complexes with 5 or fewer members (96 pairs). In this case, the median  $\rho$  value increase is 15.8% (WMSRT  $p = 6.2\text{E-}03$ ). The results support the view that proteins within known complexes show higher ERCs than between complexes, and further implicate physical contact as a contributor to ERCs. However, other studies have found ERCs between proteins that do not bind to each other, but are involved in shared function, such as metabolic pathways (Clark, Alani & Aquadro, 2012). Thus, future research is needed to better understand the different biological drivers of ERCs between proteins.

## Discussion

An overwhelmingly strong pattern is an association between ACE2, its partners, and the proteins involved with coagulation, cytokine signaling, and immunity. For coagulation, this is exemplified by the enrichment for terms related to coagulation pathways in the CRR network, and the presence of the three proteins that form fibrinogen (FGA, FGB, FGG) which constitutes the clotting molecule fibrin. Abnormal clotting and coagulation such as “hypercoagulability” has been observed as a major symptom of COVID-19 infection (Fei et al., 2020). Additionally, disseminated intravascular coagulation (DIC) due to COVID-19 has been found more frequently in fatal cases of COVID-19 than non-fatal cases (Seitz & Schramm, 2020). Levi et al. (2020) have noted that low-grade DIC often seen in COVID-19 is associated with a sudden decrease in plasma fibrinogen before death. This makes the connection with the various fibrinogen subcomponents even more striking. Our network data suggest that ACE2’s connection to fibrinogen is mediated through Clusterin and GPR141 (Fig. 3). The chaperone protein Clusterin’s role in removing misfolded proteins in the blood and its common association with fibrinogen in blood plasma (Wyatt & Wilson, 2010) lend credence to these ERC findings. What remains unclear is the nature of potential functional interactions between ACE2 and Clusterin, but the ERC results suggest that this warrants further attention. The discovery of a strong ERC association of Clusterin and GPR141 is a novel finding, as functional information on GPR141 is

largely lacking. ERC analysis indicates that these proteins functionally interact, likely involving coagulation processes.

Another mechanism for ACE2's influence on the coagulation effects of COVID-19, based on ERCs, is through F5. F5 canonically is activated by the same enzyme (Thrombin) that converts fibrinogen into fibrin for clotting (Omarova et al., 2013). Omarova et al. (2013) further report that inhibition of F5 can enhance an anticoagulant ability of an alternate fibrinogen that utilizes a different isoform of FGG, fibrinogen  $\gamma'$ . Thus, we hypothesize that abnormal coagulation activity may (in part) be driven by disruptions in ACE2-F5 protein interactions, which could reduce anticoagulant feedback mechanisms. F5 is also found to have many significant ERCs outside of the coagulation pathway, connecting to various immunity-related pathways (Fig. 4, Supplementary File 1). The ERC results for GPR141 and F5 reveal how ERC analysis may be useful in providing testable hypotheses for functions of understudied proteins, and to investigate additional functional roles on well-studied proteins.

A second major finding is ACE2 protein-protein interactions that connect to cytokine signaling and immunity. "Cytokine storms", an overreaction of the immune system which can lead to inflammation and organ failure, is a second major hallmark of severe COVID-19, and its management is a major target of medical treatment research (Luo et al., 2020; Mangalmurti & Hunter, 2020). Chemokines are a class of cytokines that act as immune cell attractants (Coperchini et al., 2020), and an increase in chemokine production may be characteristic of COVID-19 infection (Coperchini et al., 2020). XCR1 is a receptor of XCL1 chemokines, mostly expressed in dendritic cells, and plays a role in cytotoxic immune responses (Lei & Takahama, 2012). The XCR1 protein, strikingly, is the second-highest ERC to ACE2 and has already been implicated in severe COVID-19 infection (Severe Covid-19 GWAS Group, 2020). While the specific mechanism by which XCR1 might play a role in severe COVID-19 is not yet known, ERC results indicate its role may be mediated by ACE2 with XCR1's ERCs also possibly indicating a broader functional role in coagulation. Excessive Inflammatory response, particularly as a consequence of cytokine storms, is a clear pathology of COVID-19.

Type 1 interferons are among the first types of cytokines produced after viral infection (García-Sastre & Biron, 2006; Sallard et al., 2020). A component of the type 1 interferon receptor, IFNAR2, is among the strongest ACE2 ERCs, possibly linking ACE2 to the type 1 interferon immunity response. Notably, IFNAR2 has been implicated in severe COVID-19 infection (Pairo-Castineira et al., 2021). Since type 1 interferons have shown some initial efficacy in treating COVID-19 infection (Sallard et al., 2020), it is possible that the SARS-CoV-2 virus interaction with both receptor and soluble ACE2 interferes with type 1 interferon response, as low levels of type 1 interferons have been found in COVID-19 patients (Salman et al., 2021). Another connection of ACE2 with immunity may be mediated by the toll-like receptor TLR8 (a strong ACE2 ERC), among TLRs believed to regulate platelet circulation in response to inflammation (Beaulieu & Freedman, 2010) providing possible avenues for interaction with soluble ACE2 in

blood. Genetic variants in TLRs (including TLR8) may affect COVID-19 susceptibility (Lee, Lee & Kong, 2020). Thus, there are many potential avenues for ACE2 protein interactions contributing to immune dysregulation in COVID-19 disease, which may warrant further investigation given the strong ERC associations of ACE2 with proteins relevant to immunity, although the functional bases of such interactions are unknown. Other ACE2 network ERCs of interest are relevant to kidney disease, cardiovascular disease, male fertility, Alzheimer's disease, and DNA damage checkpoint signaling. These are discussed further in the Supplementary Text.

Overall, the underlying concept behind the evolutionary rate correlation approach (also called evolutionary rate covariance or evolutionary rate coevolution) is that coevolving proteins will show correlated rates of change across evolution and that this reflects functional interactions (Clark, Alani & Aquadro, 2012; Wolfe & Clark, 2015). Clark and colleagues have developed a web interface ([https://csb.pitt.edu/erc\\_analysis/](https://csb.pitt.edu/erc_analysis/)) to screen for ERC interactions for *Drosophila*, yeast, and mammals. Their mammalian data set is based on 33 mammalian species (Priedigkeit, Wolfe & Clark, 2015; Wolfe & Clark, 2015). We have compared their output for ACE2 to our analyses and found only one overlapping protein (XCR1) between their significant ERCs ( $p < 0.05$ ) and our top 2% ACE2 ERCs. There are many methodological differences between our approaches, including the number and specific mammalian taxa used, the method for calculating protein rates, and the phylogeny used for calculating branch lengths. In addition, their dataset includes 17,487 proteins, whereas our analysis is currently restricted to 1,953 proteins for which we were confident about 1:1 orthology and therefore for which there are minimal paralogy complications. Furthermore, we are uncertain how their database dealt with potential short branch artifacts on ERC calculations. In our case, we found that short branches in the phylogeny resulted in significant correlations between branch time and protein rate, thus both inflating estimated ERCs and introducing branch time as a confounding factor which can lead to spurious correlations, and we removed these by branch trimming.

In another study, Braun et al. (2020) applied a “phylogenetic profiling” approach to identify ACE2 interacting proteins relevant to possible drug targets for COVID-19. Phylogenetic profiling generally screens multiple genomes for presence-absence correlations of protein combinations, as a method to detect candidate protein interactions (Pellegrini et al., 1999). However, Braun et al. (2020) use a modification of the method that also incorporates a BLAST-based distance metric from human ACE2 across taxa ranging from humans to fungi. When we focus on proteins common between our set and their mammalian data set (1,875 proteins), there are three shared proteins among the top 1% for both sets, Androgen Receptor (AR) and Angiomotin (AMOT), and nucleotide excision repair protein homolog MMS19, with no additional proteins in the respective top 2% sets. We suggest that our direct measures of protein evolutionary rates, which utilize aligned sequences and phylogenetic analysis, may be a more sensitive approach for finding evolutionary interactions among proteins in mammals. Obviously, future validation studies are needed to reveal which approaches are most effective at detecting candidate protein interactions, or whether each has its own merits for the detection of different interactions.

Experimental validations of novel ACE2 protein associations predicted by our ERC approach are clearly needed. A necessary first step is to establish whether ACE2 has binding affinities *in vitro* and *in vivo* with proteins showing high evolutionary correlation to it, in particular CLU, XCR1, GEN1, and IFNAR2. Similar binding affinity is predicted between CLU and GPR141 based on their high reciprocal rank ERCs. CLU-FGG and GPR141-FGB provide connections to fibrinogen based on their evolutionary correlations, suggesting binding affinities. Applicable methods could include protein complex immunoprecipitation, tagged protein analysis, and yeast-two-hybrid analysis (Rao et al., 2014).

We have begun preliminary analyses using short (10mer) amino acid sequences to identify predicted sites of interaction among protein partners. These data may be able to inform docking simulations for protein pairs using software that allows for the incorporation of *a priori* predicted interfaces (Van Zundert et al., 2016; Pagadala, Syed & Tuszyński, 2017). For example, these 10mer analyses can be used to determine likely regions of binding affinity between ACE2 and Clusterin, for experimental validation through mutational analysis. Similarly, coagulation factor V shows high ERCs for non-canonical proteins, which can be investigated to determine whether F5 has novel functions outside of the coagulation pathway.

## Conclusions

In this paper, we take an exploratory approach to ACE2 protein interactions using evolutionary rate correlations. Our key findings are that the ERC analysis predicts ACE2 to have previously unidentified protein partners, and to be part of interaction networks relevant to COVID-19 pathologies. Most notably, ACE2 forms strong ERC networks relevant to coagulation and immunity. A potential mechanism is that reduced abundance of membrane-bound ACE2 disrupts signaling networks. Additionally, the presence of the soluble ACE2 ectodomain may explain the systemic pathologies of COVID-19 infection as its circulation in the blood can affect pathways throughout the body. We recognize that the new ACE2 protein connections predicted by ERCs may not be causal in severe COVID-19 pathologies. However, our novel findings that the ACE2 ERC network connects to coagulation and immunity pathways is noteworthy, with clear potential implications to some of the unusual features of COVID-19. In addition, results may have relevance to other functions of ACE2, such as circulatory homeostasis and digestion. The ERC analysis predicts additional protein connections that can be relevant to biological processes and disease. For instance, ERCs predict novel interactions for cytokine and immunity related proteins, such as for XCR1, IFNLR1, IFNAR2, and TLR8. Future investigations of the ERC networks of these and related proteins could be worthwhile. ERCs also suggest strong but previously undescribed connections for proteins, such as CLU, GPR141, F5, and GEN1. Validation studies are necessary to determine to what extent strong ERCs predict biological interactions among proteins, such as the ones detected here.

Further computational analyses of ERCs are needed to better understand their relationship to protein function and evolution. For instance, machine learning and simulation approaches can be used to determine which aspects of protein structure, amino acid properties, and rates of protein evolution, improve ERC predictive power. We are currently expanding the mammalian protein set for such analyses. Finally, if evidence mounts that ERCs can be informative in predicting protein interactions, the approach can be applied more broadly as an additional tool for detecting protein interaction networks involved in many biological processes and disease.

## Acknowledgements

We thank Zhichao Yan for guidance with the mammalian protein database and discussions on ERC methods and independent contrasts. Also thanked are J. Fay, J. Fry, J. Jaenike, A. Kingsley, A. Larracuente, E. Sia, S. Ghaemmaghami for discussions and helpful feedback, and M. Tsuchiya for support and helpful discussions. Special thanks also go to Nathaniel and Helen Wisch for their support.

## References

Aase K, Ernvist M, Ebarasi L, Jakobsson L, Majumdar A, Yi C, Birot O, Ming Y, Kvanta A, Edholm D, Aspenström P, Kissil J, Claesson-Welsh L, Shimono A, Holmgren L. 2007. Angiomotin regulates endothelial cell migration during embryonic angiogenesis. *Genes and Development* 21:2055–2068. DOI: 10.1101/gad.432007.

Bastolla U. 2020. Mathematical model of SARS-CoV-2 propagation versus ACE2 fits COVID-19 lethality across age and sex and predicts that of SARS, supporting possible therapy. *arXiv Preprint*.

Bastolla U, Chambers P, Abia D, García-Bermejo M-L, Fresno M. 2021. Is Covid-19 severity associated with ACE2 degradation? *arXiv Preprint*.

Beaulieu LM, Freedman JE. 2010. The role of inflammation in regulating platelet production and function: Toll-like receptors in platelets and megakaryocytes. *Thrombosis Research* 125:205–209. DOI: 10.1016/j.thromres.2009.11.004.

Bernas SN, Baldauf H, Wendler S, Heidenreich F, Lange V, Hofmann JA, Sauter J, Schmidt AH, Schetelig J. 2021. CCR5Δ32 mutations do not determine COVID-19 disease course. *International Journal of Infectious Diseases* 105:653–655. DOI: 10.1016/j.ijid.2021.02.108.

Böhm S, Szakal B, Herken BW, Sullivan MR, Mihalevic MJ, Kabbinavar FF, Branzei D, Clark NL, Bernstein KA. 2016. The budding yeast ubiquitin protease Ubp7 is a novel component involved in s phase progression. *Journal of Biological Chemistry* 291:4442–4452. DOI: 10.1074/jbc.M115.671057.

Bonaventura A, Vecchié A, Dagna L, Martinod K, Dixon DL, Van Tassell BW, Dentali F, Montecucco F, Massberg S, Levi M, Abbate A. 2021. Endothelial dysfunction and immunothrombosis as key pathogenic mechanisms in COVID-19. *Nature Reviews Immunology* 21:319–329. DOI: 10.1038/s41577-021-00536-9.

Bouaziz JD, Duong TA, Jachiet M, Velter C, Lestang P, Cassius C, Arsouze A, Domergue Than Trong E, Bagot M, Begon E, Sulimovic L, Rybojad M. 2020. Vascular skin symptoms in COVID-19: a French observational study. *Journal of the European Academy of Dermatology and Venereology* 34:e451–e452. DOI: 10.1111/jdv.16544.

Bratt A, Birot O, Sinha I, Veitonmäki N, Aase K, Ernvist M, Holmgren L. 2005. Angiomotin regulates endothelial cell-cell junctions and cell motility. *Journal of Biological Chemistry* 280:34859–34869. DOI:

10.1074/jbc.M503915200.

Braun M, Sharon E, Unterman I, Miller M, Shtern AM, Benenson S, Vainstein A, Tabach Y. 2020. ACE2 Co-evolutionary Pattern Suggests Targets for Pharmaceutical Intervention in the COVID-19 Pandemic. *iScience* 23:101384. DOI: 10.1016/j.isci.2020.101384.

Brunette GJ, Jamalruddin MA, Baldock RA, Clark NL, Bernstein KA. 2019. Evolution-based screening enables genome-wide prioritization and discovery of DNA repair genes. *Proceedings of the National Academy of Sciences of the United States of America* 116:19593–19599. DOI: 10.1073/pnas.1906559116.

Burrell LM, Johnston CI, Tikellis C, Cooper ME. 2004. ACE2, a new regulator of the renin-angiotensin system. *Trends in Endocrinology and Metabolism* 15:166–169. DOI: 10.1016/j.tem.2004.03.001.

Busino L, Bassermann F, Maiolica A, Lee C, Nolan PM, Godinho SIH, Draetta GF, Pagano M. 2007. SCFFbxl3 controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins. *Science* 316:900–904. DOI: 10.1126/science.1141194.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. DOI: 10.1093/bioinformatics/btp348.

Chan YW, West S. 2015. GEN1 promotes Holliday junction resolution by a coordinated nick and counter-nick mechanism. *Nucleic Acids Research* 43:10882–10892. DOI: 10.1093/nar/gkv1207.

Chen ZG, Wang ZN, Yan Y, Liu J, He TT, Thong KT, Ong YK, Chow VTK, Tan K Sen, Wang DY. 2019. Upregulation of cell-surface mucin MUC15 in human nasal epithelial cells upon influenza A virus infection. *BMC Infectious Diseases* 19:622. DOI: 10.1186/s12879-019-4213-y.

Chen C, Zhang XR, Ju ZY, He WF. 2020. Advances in the research of mechanism and related immunotherapy on the cytokine storm induced by coronavirus disease 2019. *Zhonghua shao shang za zhi = Zhonghua shaoshang zazhi = Chinese journal of burns* 36:471–475. DOI: 10.3760/cma.j.cn501120-20200224-00088.

Clark NL, Alani E, Aquadro CF. 2012. Evolutionary rate covariation reveals shared functionality and coexpression of genes. *Genome Research* 22:714–720. DOI: 10.1101/gr.132647.111.

Colgren J, Nichols SA. 2019. Evolution as a guide for experimental cell biology. *PLoS Genetics* 15:e1007937. DOI: 10.1371/journal.pgen.1007937.

Connors JM, Levy JH. 2020. COVID-19 and its implications for thrombosis and anticoagulation. *Blood* 135:2033–2040. DOI: 10.1182/BLOOD.2020006000.

Coperchini F, Chiovato L, Croce L, Magri F, Rotondi M. 2020. The cytokine storm in COVID-19: An overview of the involvement of the chemokine/chemokine-receptor system. *Cytokine and Growth Factor Reviews* 53:25–32. DOI: 10.1016/j.cytogfr.2020.05.003.

Deng Q, Rasool R ur, Russell RM, Natesan R, Asangani IA. 2021. Targeting androgen regulation of TMPRSS2 and ACE2 as a therapeutic strategy to combat COVID-19. *iScience* 24:102254. DOI: 10.1016/j.isci.2021.102254.

Dolan ME, Hill DP, Mukherjee G, McAndrews MS, Chesler EJ, Blake JA. 2020. Investigation of COVID-19 comorbidities reveals genes and pathways coincident with the SARS-CoV-2 viral disease. *Scientific Reports* 10:1–11. DOI: 10.1038/s41598-020-77632-8.

Eisenreich A, Orphal M, Böhme K, Kreutz R. 2020. Tmem63c is a potential pro-survival factor in angiotensin II-treated human podocytes. *Life Sciences* 258. DOI: 10.1016/j.lfs.2020.118175.

Fang Z, Wolf A, Liao Y, McKay A, Fröhlich F, Kimmel J, Xiaohui L, sorrg. 2021. zqfang/GSEAp: gseapy-v0.10.3. DOI: 10.5281/ZENODO.4553090.

Fei Y, Tang N, Liu H, Cao W. 2020. Coagulation dysfunction: A hallmark in COVID-19. *Archives of Pathology and*

*Laboratory Medicine* 144:1223–1229. DOI: 10.5858/arpa.2020-0324-SA.

Fina L, Molgaard H, Robertson D, Bradley N, Monaghan P, Delia D, Sutherland D, Baker M, Greaves M. 1990. Expression of the CD34 gene in vascular endothelial cells. *Blood* 75:2417–2426. DOI: 10.1182/blood.v75.12.2417.2417.

Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF. 2014. Evolutionary Rate Covariation Identifies New Members of a Protein Network Required for *Drosophila melanogaster* Female Post-Mating Responses. *PLoS Genetics* 10:e1004108. DOI: 10.1371/journal.pgen.1004108.

Forsbach A, Samulowitz U, Völz K, Hofmann HP, Noll B, Tluk S, Schmitz C, Wader T, Müller C, Podzuweit A, Lohner A, Curdt R, Uhlmann E, Vollmer J. 2011. Dual or triple activation of TLR7, TLR8, and/or TLR9 by single-stranded oligoribonucleotides. *Nucleic Acid Therapeutics* 21:423–436. DOI: 10.1089/nat.2011.0323.

Fricke-Galindo I, Falfán-Valencia R. 2021. Genetics Insight for COVID-19 Susceptibility and Severity: A Review. *Frontiers in Immunology* 12. DOI: 10.3389/fimmu.2021.622176.

García-Sastre A, Biron CA. 2006. Type 1 interferons and the virus-host relationship: A lesson in détente. *Science* 312:879–882. DOI: 10.1126/science.1125676.

Garland T, Harvey PH, Ives AR. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 41:18–32. DOI: 10.1093/sysbio/41.1.18.

Giurgiu M, Reinhard J, Brauner B, Dunger-Kaltenbach I, Fobo G, Frishman G, Montrone C, Ruepp A. 2019. CORUM: The comprehensive resource of mammalian protein complexes - 2019. *Nucleic Acids Research* 47:D559–D563. DOI: 10.1093/nar/gky973.

Godin SK, Meslin C, Kabbinavar F, Bratton-Palmer DS, Hornack C, Mihalevic MJ, Yoshida K, Sullivan M, Clark NL, Bernstein KA. 2015. Evolutionary and functional analysis of the invariant SWIM domain in the conserved Shu2/SWS1 protein family from *Saccharomyces cerevisiae* to *Homo sapiens*. *Genetics* 199:1023–1033. DOI: 10.1534/genetics.114.173518.

Gómez J, Cuesta-Llavona E, Albaiceta GM, García-Clemente M, López-Larrea C, Amado-Rodríguez L, López-Alonso I, Hermida T, Enríquez AI, Gil H, Alonso B, Iglesias S, Suárez-Alvarez B, Alvarez V, Coto E. 2020. The CCR5-delta32 variant might explain part of the association between COVID-19 and the chemokine-receptor gene cluster. *medRxiv*:2020.11.02.20224659. DOI: 10.1101/2020.11.02.20224659.

Gotluru C, Roach A, Cherry SH, Runowicz CD. 2021. Sex, Hormones, Immune Functions, and Susceptibility to Coronavirus Disease 2019 (COVID-19)-Related Morbidity. *Obstetrics and gynecology* 137:423–429. DOI: 10.1097/AOG.0000000000004275.

Gupta A, Madhavan M V., Sehgal K, Nair N, Mahajan S, Sehrawat TS, Bikdeli B, Ahluwalia N, Ausiello JC, Wan EY, Freedberg DE, Kirtane AJ, Parikh SA, Maurer MS, Nordvig AS, Accili D, Bathon JM, Mohan S, Bauer KA, Leon MB, Krumholz HM, Uriel N, Mehra MR, Elkind MSV, Stone GW, Schwartz A, Ho DD, Bilezikian JP, Landry DW. 2020. Extrapulmonary manifestations of COVID-19. *Nature Medicine* 26:1017–1032. DOI: 10.1038/s41591-020-0968-3.

Hanson A, Cohen H, Wang H, Shekhar N, Shah C, Dhaneshwar A, Harvey BW, Murray R, Harvey CJ. 2020. Impaired ICOS signaling between Tfh and B cells distinguishes hospitalized from ambulatory CoViD-19 patients. *medRxiv*:2020.12.16.20248343. DOI: 10.1101/2020.12.16.20248343.

Hubacek JA, Dusek L, Majek O, Adamek V, Cervinkova T, Dlouha D, Pavel J, Adamkova V. 2021. CCR5Δ32 Deletion as a Protective Factor in Czech First-Wave COVID-19 Subjects. *Physiological Research* 70:111–115. DOI: 10.33549/physiolres.934647.

Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution* 33:1635–1638. DOI: 10.1093/molbev/msw046.

Hutloff A, Dittrich AM, Beier KC, Eljaschewitsch B, Kraft R, Anagnostopoulos I, Kroczek RA. 1999. ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. *Nature* 397:263–266. DOI: 10.1038/16717.

Itakura E, Chiba M, Murata T, Matsuura A. 2020. Heparan sulfate is a clearance receptor for aberrant extracellular proteins. *Journal of Cell Biology* 219. DOI: 10.1083/JCB.201911126.

Jain R, Ramaswamy S, Harilal D, Uddin M, Loney T, Nowotny N, Alsuwaidi H, Varghese R, Deesi Z, Alkhajeh A, Khansaheb H, Alsheikh-Ali A, Abou Tayoun A. 2021. Host transcriptomic profiling of COVID-19 patients with mild, moderate, and severe clinical outcomes. *Computational and Structural Biotechnology Journal* 19:153–160. DOI: 10.1016/j.csbj.2020.12.016.

Jeffery CJ. 1999. Moonlighting proteins. *Trends in Biochemical Sciences* 24:8–11. DOI: 10.1016/S0968-0004(98)01335-8.

Jin HY, Chen LJ, Zhang ZZ, Xu Y Le, Song B, Xu R, Oudit GY, Gao PJ, Zhu DL, Zhong JC. 2015. Deletion of angiotensin-converting enzyme 2 exacerbates renal inflammation and injury in apolipoprotein E-deficient mice through modulation of the nephrin and TNF-alpha-TNFRSF1A signaling. *Journal of Translational Medicine* 13:1–16. DOI: 10.1186/s12967-015-0616-8.

Jones VG, Mills M, Suarez D, Hogan CA, Yeh D, Bradley Segal J, Nguyen EL, Barsh GR, Maskatia S, Mathew R. 2020. COVID-19 and Kawasaki Disease: Novel Virus and Novel Case. *Hospital pediatrics* 10:537–540. DOI: 10.1542/hpeds.2020-0123.

De Juan D, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. *Nature Reviews Genetics* 14:249–261. DOI: 10.1038/nrg3414.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28:27–30. DOI: 10.1093/nar/28.1.27.

Kasela S, Daniloski Z, Jordan TX, Tenover BR, Sanjana NE, Lappalainen T. 2021. Integrative approach identifies SLC6A20 and CXCR6 as putative causal genes for the COVID-19 GWAS signal in the 3p21.31 locus. *medRxiv*:2021.04.09.21255184. DOI: 10.1101/2021.04.09.21255184.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780. DOI: 10.1093/molbev/mst010.

Khalili MA, Leisegang K, Majzoub A, Finelli R, Selvam MKP, Henkel R, Mojgan M, Agarwal A. 2020. Male fertility and the COVID-19 pandemic: Systematic review of the literature. *World Journal of Men's Health* 38:1–15. DOI: 10.5534/WJMH.200134.

Kowalczyk A, Gbadamosi O, Kolor K, Sosa J, Croix CS, Gibson G, Chikina M, Aizenman E, Clark N, Kiselyov K. 2021. Evolutionary rate covariation identifies SLC30A9 (ZnT9) as a mitochondrial zinc transporter. *bioRxiv*:2021.04.22.440839. DOI: 10.1101/2021.04.22.440839.

Kriventseva E V., Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* 47:D807–D811. DOI: 10.1093/nar/gky1053.

Kriventseva E V., Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM. 2015. OrthoDB v8: Update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Research* 43:D250–D256. DOI: 10.1093/nar/gku1220.

Kuba K, Imai Y, Ohto-Nakanishi T, Penninger JM. 2010. Trilogy of ACE2: A peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. *Pharmacology and Therapeutics* 128:119–128. DOI: 10.1016/j.pharmthera.2010.06.003.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular biology and evolution* 34:1812–1819. DOI: 10.1093/molbev/msx116.

Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581:215–220. DOI: 10.1038/s41586-020-2180-5.

Lazzaroni MG, Piantoni S, Masneri S, Garrafa E, Martini G, Tincani A, Andreoli L, Franceschini F. 2021. Coagulation dysfunction in COVID-19: The interplay between inflammation, viral infection and the coagulation system. *Blood Reviews* 46:100745. DOI: 10.1016/j.blre.2020.100745.

Lee IH, Lee JW, Kong SW. 2020. A survey of genetic variants in SARS-CoV-2 interacting domains of ACE2, TMPRSS2 and TLR3/7/8 across populations. *Infection, Genetics and Evolution* 85:104507. DOI: 10.1016/j.meegid.2020.104507.

Leenaerts D, Loyau S, Mertens JC, Boisseau W, Michel JB, Lambeir AM, Jandrot-Perrus M, Hendriks D. 2018. Carboxypeptidase U (CPU, carboxypeptidase B2, activated thrombin-activatable fibrinolysis inhibitor) inhibition stimulates the fibrinolytic rate in different in vitro models. *Journal of Thrombosis and Haemostasis* 16:2057–2069. DOI: 10.1111/jth.14249.

Lei Y, Takahama Y. 2012. XCL1 and XCR1 in the immune system. *Microbes and Infection* 14:262–267. DOI: 10.1016/j.micinf.2011.10.003.

Leung LLK, Morser J. 2018. Carboxypeptidase B2 and carboxypeptidase N in the crosstalk between coagulation, thrombosis, inflammation, and innate immunity. *Journal of Thrombosis and Haemostasis* 16:1474–1486. DOI: 10.1111/jth.14199.

Levi M, Thachil J, Iba T, Levy JH. 2020. Coagulation abnormalities and thrombosis in patients with COVID-19. *The Lancet Haematology* 7:e438–e440. DOI: 10.1016/S2352-3026(20)30145-9.

van Lier D, Kox M, Santos K, van der Hoeven H, Pillay J, Pickkers P. 2021. Increased blood angiotensin converting enzyme 2 activity in critically ill COVID-19 patients. *ERJ Open Research* 7:00848–02020. DOI: 10.1183/23120541.00848-2020.

Liu D, Yang J, Feng B, Lu W, Zhao C, Li L. 2021. Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19. *Journal of Infection* 82:126–132. DOI: 10.1016/j.jinf.2020.11.031.

López-León S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, Villapol S. 2021. More than 50 Long-Term Effects of COVID-19: A Systematic Review and Meta-Analysis. *SSRN Electronic Journal*:2021.01.27.21250617. DOI: 10.2139/ssrn.3769978.

Luo P, Liu Y, Qiu L, Liu X, Liu D, Li J. 2020. Tocilizumab treatment in COVID-19: A single center experience. *Journal of Medical Virology* 92:814–818. DOI: 10.1002/jmv.25801.

Luo S, Zhang X, Xu H. 2020. Don't Overlook Digestive Symptoms in Patients With 2019 Novel Coronavirus Disease (COVID-19). *Clinical Gastroenterology and Hepatology* 18:1636–1637. DOI: 10.1016/j.cgh.2020.03.043.

Magro C, Mulvey JJ, Berlin D, Nuovo G, Salvatore S, Harp J, Baxter-Stoltzfus A, Laurence J. 2020. Complement associated microvascular injury and thrombosis in the pathogenesis of severe COVID-19 infection: A report of five cases. *Translational Research* 220:1–13. DOI: 10.1016/j.trsl.2020.04.007.

Mamoor S. 2020. The transcription factor ZBTB43 is differentially expressed and transcriptionally induced in models of coronavirus infection. DOI: 10.31219/osf.io/jhmfv.

Mangalmurti N, Hunter CA. 2020. Cytokine Storms: Understanding COVID-19. *Immunity* 53:19–25. DOI:

10.1016/j.immuni.2020.06.017.

Matsumoto T, Shiina H, Kawano H, Sato T, Kato S. 2008. Androgen receptor functions in male and female physiology. *Journal of Steroid Biochemistry and Molecular Biology* 109:236–241. DOI: 10.1016/j.jsbmb.2008.03.023.

McKechnie JL, Blish CA. 2020. The Innate Immune System: Fighting on the Front Lines or Fanning the Flames of COVID-19? *Cell Host and Microbe* 27:863–869. DOI: 10.1016/j.chom.2020.05.009.

Medcalf RL, Keragala CB, Myles PS. 2020. Fibrinolysis and COVID-19: A plasmin paradox. *Journal of Thrombosis and Haemostasis* 18:2118–2122. DOI: 10.1111/jth.14960.

Medina-Enríquez MM, Lopez-León S, Carlos-Escalante JA, Aponte-Torres Z, Cuapio A, Wegman-Ostrosky T. 2020. ACE2: the molecular doorway to SARS-CoV-2. *Cell and Bioscience* 10:1–17. DOI: 10.1186/s13578-020-00519-8.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R, Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* 37:1530–1534. DOI: 10.1093/molbev/msaa015.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M, Izquierdo-Carrasco F, Wappler T, Rust J, Aberer AJ, Aspöck U, Aspöck H, Bartel D, Blanke A, Berger S, Böhm A, Buckley TR, Calcott B, Chen J, Friedrich F, Fukui M, Fujita M, Greve C, Grobe P, Gu S, Huang Y, Jermiin LS, Kawahara AY, Krogmann L, Kubiak M, Lanfear R, Letsch H, Li Y, Li Z, Li J, Lu H, Machida R, Mashimo Y, Kapli P, McKenna DD, Meng G, Nakagaki Y, Navarrete-Heredia JL, Ott M, Ou Y, Pass G, Podsiadlowski L, Pohl H, Von Reumont BM, Schütte K, Sekiya K, Shimizu S, Slipinski A, Stamatakis A, Song W, Su X, Szucsich NU, Tan M, Tan X, Tang M, Tang J, Timelthaler G, Tomizuka S, Trautwein M, Tong X, Uchifune T, Walzl MG, Wiegmann BM, Wilbrandt J, Wipfler B, Wong TKF, Wu Q, Wu G, Xie Y, Yang S, Yang Q, Yeates DK, Yoshizawa K, Zhang Q, Zhang R, Zhang W, Zhang Y, Zhao J, Zhou C, Zhou L, Ziesmann T, Zou S, Li Y, Xu X, Zhang Y, Yang H, Wang J, Wang J, Kjer KM, Zhou X. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346:763–767. DOI: 10.1126/science.1257570.

Morand A, Urbina D, Fabre A. 2020. COVID-19 and Kawasaki like disease : the known-known , the unknown-known and the unknown-unknown. *Preprints*:2020050160. DOI: 10.20944/PREPRINTS202005.0160.V1.

Morser J, Shao Z, Nishimura T, Zhou Q, Zhao L, Higgins J, Leung LLK. 2018. Carboxypeptidase B2 and N play different roles in regulation of activated complements C3a and C5a in mice. *Journal of Thrombosis and Haemostasis* 16:991–1002. DOI: 10.1111/jth.13964.

Mosesson MW. 2005. Fibrinogen and fibrin structure and functions. In: *Journal of Thrombosis and Haemostasis*. John Wiley & Sons, Ltd, 1894–1904. DOI: 10.1111/j.1538-7836.2005.01365.x.

Myllyharju J, Kivirikko KI. 2001. Collagens and collagen-related diseases. *Annals of Medicine* 33:7–21. DOI: 10.3109/07853890109002055.

Niazkar HR, Zibaei B, Nasimi A, Bahri N. 2020. The neurological manifestations of COVID-19: a review article. *Neurological Sciences* 41:1667–1671. DOI: 10.1007/s10072-020-04486-3.

Omarova F, Uitte De Willige S, Ariëns RAS, Rosing J, Bertina RM, Castoldi E. 2013. Inhibition of thrombin-mediated factor V activation contributes to the anticoagulant activity of fibrinogen  $\gamma'$ . *Journal of Thrombosis and Haemostasis* 11:1669–1678. DOI: 10.1111/jth.12354.

Pagadala NS, Syed K, Tuszyński J. 2017. Software for molecular docking: a review. *Biophysical Reviews* 9:91–102. DOI: 10.1007/s12551-016-0247-1.

Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH,

Russell CD, Furniss J, Richmond A, Gountouna E, Wrobel N, Harrison D, Wang B, Wu Y, Meynert A, Griffiths F, Oosthuizen W, Kousathanas A, Moutsianas L, Yang Z, Zhai R, Zheng C, Grimes G, Beale R, Millar J, Shih B, Keating S, Zechner M, Haley C, Porteous DJ, Hayward C, Yang J, Knight J, Summers C, Shankar-Hari M, Klenerman P, Turtle L, Ho A, Moore SC, Hinds C, Horby P, Nichol A, Maslove D, Ling L, McAuley D, Montgomery H, Walsh T, Pereira AC, Renieri A, Shen X, Ponting CP, Fawkes A, Tenesa A, Caulfield M, Scott R, Rowan K, Murphy L, Openshaw PJM, Semple MG, Law A, Vitart V, Wilson JF, Baillie JK. 2021. Genetic mechanisms of critical illness in COVID-19. *Nature* 591:92–98. DOI: 10.1038/s41586-020-03065-y.

Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. 1999. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proceedings of the National Academy of Sciences of the United States of America* 96:4285–4288. DOI: 10.1073/pnas.96.8.4285.

Peng X, Xu E, Liang W, Pei X, Chen D, Zheng D, Zhang Y, Zheng C, Wang P, She S, Zhang Y, Ma J, Mo X, Zhang Y, Ma D, Wang Y. 2016. Identification of FAM3D as a new endogenous chemotaxis agonist for the formyl peptide receptors. *Journal of Cell Science* 129:1831–1842. DOI: 10.1242/jcs.183053.

Priedigkeit N, Wolfe N, Clark NL. 2015. Evolutionary Signatures amongst Disease Genes Permit Novel Methods for Gene Prioritization and Construction of Informative Gene-Based Networks. *PLoS Genetics* 11:1–17. DOI: 10.1371/journal.pgen.1004967.

Rabb H. 2020. Kidney diseases in the time of COVID-19: Major challenges to patient care. *Journal of Clinical Investigation* 130:2749–2751. DOI: 10.1172/JCI138871.

Rao VS, Srinivas K, Sujini GN, Kumar GNS. 2014. Protein-Protein Interaction Detection: Methods and Analysis. *International Journal of Proteomics* 2014:1–12. DOI: 10.1155/2014/147648.

Raza Q, Choi JY, Li Y, O'Dowd RM, Watkins SC, Chikina M, Hong Y, Clark NL, Kwiatkowski A V. 2019. Evolutionary rate covariation analysis of E-cadherin identifies Raskol as a regulator of cell adhesion and actin dynamics in Drosophila. *PLoS Genetics* 15:e1007720. DOI: 10.1371/journal.pgen.1007720.

Reilly DF, Westgate EJ, FitzGerald GA. 2007. Peripheral circadian clocks in the vasculature. *Arteriosclerosis, Thrombosis, and Vascular Biology* 27:1694–1705. DOI: 10.1161/ATVBAHA.107.144923.

Sallard E, Lescure FX, Yazdanpanah Y, Mentre F, Peiffer-Smadja N. 2020. Type 1 interferons as a potential treatment against COVID-19. *Antiviral Research* 178:104791. DOI: 10.1016/j.antiviral.2020.104791.

Salman AA, Waheed MH, Ali-Abdulsahib AA, Atwan ZW. 2021. Low type i interferon response in covid-19 patients: Interferon response may be a potential treatment for covid-19. *Biomedical Reports* 14:1–5. DOI: 10.3892/br.2021.1419.

Samavati L, Uhal BD. 2020. ACE2, Much More Than Just a Receptor for SARS-CoV-2. *Frontiers in Cellular and Infection Microbiology* 10:317. DOI: 10.3389/fcimb.2020.00317.

Samuel RM, Majd H, Richter MN, Ghazizadeh Z, Zekavat SM, Navickas A, Ramirez JT, Asgharian H, Simoneau CR, Bonser LR, Koh KD, Garcia-Knight M, Tassetto M, Sunshine S, Farahvashi S, Kalantari A, Liu W, Andino R, Zhao H, Natarajan P, Erle DJ, Ott M, Goodarzi H, Fattah F. 2020. Androgen Signaling Regulates SARS-CoV-2 Receptor Levels and Is Associated with Severe COVID-19 Symptoms in Men. *Cell Stem Cell* 27:876–889.e12. DOI: 10.1016/j.stem.2020.11.009.

Sánchez-Martín P, Komatsu M. 2020. Heparan sulfate and clusterin: Cleaning squad for extracellular protein degradation. *Journal of Cell Biology* 219. DOI: 10.1083/JCB.202001159.

Seabold S, Perktold J. 2010. Statsmodels: Econometric and Statistical Modeling with Python. In: *Proceedings of the 9th Python in Science Conference*. 92–96. DOI: 10.25080/majora-92bf1922-011.

Seitz R, Schramm W. 2020. DIC in COVID-19: Implications for prognosis and treatment? *Journal of Thrombosis*

and *Haemostasis* 18:1798–1799. DOI: 10.1111/jth.14878.

Severe Covid-19 GWAS Group. 2020. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *New England Journal of Medicine* 383:1522–1534. DOI: 10.1056/nejmoa2020283.

Siddiqi HK, Libby P, Ridker PM. 2021. COVID-19 – A vascular disease. *Trends in Cardiovascular Medicine* 31:1–5. DOI: 10.1016/j.tcm.2020.10.005.

Singh MK, Mobeen A, Chandra A, Joshi S, Ramachandran S. 2021. A meta-analysis of comorbidities in COVID-19: Which diseases increase the susceptibility of SARS-CoV-2 infection? *Computers in Biology and Medicine* 130:104219. DOI: 10.1016/j.combiomed.2021.104219.

Sriram K, Insel PA. 2020. A hypothesis for pathobiology and treatment of COVID-19: The centrality of ACE1/ACE2 imbalance. *British Journal of Pharmacology* 177:4825–4844. DOI: 10.1111/bph.15082.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Iny Stein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. 2016. The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics* 2016:1.30.1–1.30.33. DOI: 10.1002/cpb1.5.

Tafuri A, Shahinian A, Bladt F, Yoshinaga SK, Jordana M, Wakeham A, Boucher LM, Bouchard D, Chan VSF, Duncan G, Odermatt B, Ho A, Itie A, Horan T, Whoriskey JS, Pawson T, Penninger JM, Ohashi PS, Mak TW. 2001. ICOS is essential for effective T-helper-cell responses. *Nature* 409:105–109. DOI: 10.1038/35051113.

Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 2021. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *The Lancet Psychiatry* 8:416–427. DOI: 10.1016/s2215-0366(21)00084-5.

Terpos E, Ntanasis-Stathopoulos I, Elalamy I, Kastritis E, Sergentanis TN, Politou M, Psaltopoulou T, Gerotziafas G, Dimopoulos MA. 2020. Hematological findings and complications of COVID-19. *American Journal of Hematology* 95:834–847. DOI: 10.1002/ajh.25829.

Thomas C, Moraga I, Levin D, Krutzik PO, Podoplelova Y, Trejo A, Lee C, Yarden G, Vleck SE, Glenn JS, Nolan GP, Piehler J, Schreiber G, Garcia KC. 2011. Structural linkage between ligand discrimination and receptor activation by Type I interferons. *Cell* 146:621–632. DOI: 10.1016/j.cell.2011.06.048.

Torra R, Tazón-Vega B, Ars E, Ballarín J. 2004. Collagen type IV ( $\alpha$ 3- $\alpha$ 4) nephropathy: From isolated haematuria to renal failure. *Nephrology Dialysis Transplantation* 19:2429–2432. DOI: 10.1093/ndt/gfh435.

Uhlén M, Fagerberg L, Hallström BM, Lindsjöö C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson IM, Edlund K, Lundberg E, Navani S, Szigyarto CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, Von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, Von Heijne G, Nielsen J, Pontén F. 2015. Tissue-based map of the human proteome. *Science* 347:1260419–1260419. DOI: 10.1126/science.1260419.

Verdecchia P, Cavallini C, Spanevello A, Angeli F. 2020. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. *European Journal of Internal Medicine* 76:14–20. DOI: 10.1016/j.ejim.2020.04.037.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat I, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Vijaykumar A, Bardelli A Pietro, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik D V., Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold GL, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP,

Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmeler M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17:261–272. DOI: 10.1038/s41592-019-0686-2.

Viveiros A, Rasmussen J, Vu J, Mulvagh SL, Yip CYY, Norris CM, Oudit GY. 2021. Sex differences in COVID-19: Candidate pathways, genetics of ACE2, and sex hormones. *American Journal of Physiology - Heart and Circulatory Physiology* 320:H296–H304. DOI: 10.1152/AJPHEART.00755.2020.

Wiradjaja F, DiTommaso T, Smyth I. 2010. Basement membranes in development and disease. *Birth Defects Research Part C - Embryo Today: Reviews* 90:8–31. DOI: 10.1002/bdrc.20172.

Wolfe NW, Clark NL. 2015. ERC analysis: Web-based inference of gene function via evolutionary rate covariation. *Bioinformatics* 31:3835–3837. DOI: 10.1093/bioinformatics/btv454.

Wright FL, Vogler TO, Moore EE, Moore HB, Wohlauer M V., Urban S, Nydam TL, Moore PK, McIntyre RC. 2020. Fibrinolysis Shutdown Correlation with Thromboembolic Events in Severe COVID-19 Infection. *Journal of the American College of Surgeons* 231:193–203.e1. DOI: 10.1016/j.jamcollsurg.2020.05.007.

Wu S, Miao L, Zhou Q, Gao C, Liu J, Zhan Q, Guo B, Li F, Wang Y, Xu H, Yan H, Wu R, Zhang S, Zheng J, Yang J, Wang S, Yu W, Niu H, Li F, Yang L, Huang J, Lu X, Chen J, Tong Y, Ma L, Zhou Y, Guo Q. 2020. Suppression of Androgen Receptor (AR)-ACE2/TMPRSS2 Axis by AR Antagonists May Be Therapeutically Beneficial for Male COVID-19 Patients. *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3580526.

Wyatt AR, Wilson MR. 2010. Identification of human plasma proteins as major clients for the extracellular chaperone clusterin. *Journal of Biological Chemistry* 285:3532–3539. DOI: 10.1074/jbc.M109.079566.

Xie Z, Bailey A, Kuleshov M V., Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, Jeon M, Ma'ayan A. 2021. Gene Set Knowledge Discovery with Enrichr. *Current Protocols* 1:e90. DOI: 10.1002/cpz1.90.

Yamada T, Nakao K, Itoh H, Morii N, Shiono S, Sakamoto M, Sugawara A, Saito Y, Mukoyama M, Arai H, Eigyo M, Matsushita A, Imura H. 1988. Inhibitory action of leumorphin on vasopressin secretion in conscious rats. *Endocrinology* 122:985–990. DOI: 10.1210/endo-122-3-985.

Yan Z, Ye G, Werren JH. 2019. Evolutionary Rate Correlation between Mitochondrial-Encoded and Mitochondria-Associated Nuclear-Encoded Proteins in Insects. *Molecular Biology and Evolution* 36:1022–1036. DOI: 10.1093/molbev/msz036.

Yeung ML, Teng JLL, Jia L, Zhang C, Huang C, Cai JP, Zhou R, Chan KH, Zhao H, Zhu L, Siu KL, Fung SY, Yung S, Chan TM, To KKW, Chan JFW, Cai Z, Lau SKP, Chen Z, Jin DY, Woo PCY, Yuen KY. 2021. Soluble ACE2-mediated cell entry of SARS-CoV-2 via interaction with proteins related to the renin-angiotensin system. *Cell* 184:2212–2228.e12. DOI: 10.1016/j.cell.2021.02.053.

Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, Van Dijk M, De Vries SJ, Bonvin AMJJ. 2016. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* 428:720–725. DOI: 10.1016/j.jmb.2015.09.014.

## Supplementary Information

Supplementary Information is composed of three parts: (1) Large supplementary files deposited at FigShare (<https://doi.org/10.6084/m9.figshare.14637450>), (2) Python and R code for ERC pipelines and additional analyses deposited in GitHub (<https://github.com/austinv11/ERC-Pipeline>), and (3) Supplementary Text with embedded associated figures and tables.

---

### 1. FigShare Collection: The following files are available at

<https://doi.org/10.6084/m9.figshare.14637450>.

- **File 1:** Select proteins' 30MY ERC lists, contains multiple-test corrected p-values.
- **File 2:** Pairwise  $\rho$  and unadjusted p-value 30MY ERC matrices for all proteins.
- **File 3:** Enrichment results for select top ERC protein sets.
- **File 4:** Zip file containing the mammalian time-scaled phylogeny and maximum likelihood protein trees in newick format.
- **File 5:** Table depicting the total number of taxa present for each protein's sequence data, along with the number of taxa for which there are paralogy in the uncorrected and 30MY corrected data.
- **File 6:** Branch time to terminal branch rate correlation results for the protein set.
- **File 7:** Chi-squared test results for all proteins testing for whether there is an overrepresentation of rates below the regression line for short branches (<30MY).
- **File 8:** Branch time vs terminal branch rate residuals to branch time correlation results for the protein set.
- **File 9:** Wilcoxon matched signed-rank test significance values testing for branch adjustments following 20MY and 30MY adjustments.
- **File 10:** Coefficients for the select proteins used for the linear models containing ACE2 rate rank, Btime rate rank, and taxonomic orders as independent variables.
- **File 11:** 30MY-adjusted ERC comparisons within and between CORUM complex members.

### 2. Code Repository: <https://github.com/austinv11/ERC-Pipeline>

### 3. Supplementary Text with Embedded Figures and Tables:

Below is the supplementary text with associated figures and tables

---

**Table S1:** Index of supplementary material. (1) The FigShare Collection contains large files that are not conveniently expressed in the Supplementary Text document. (2) Code for the pipeline is made available on Github under the MIT software license. (3) All supplementary figures and the remaining tables are embedded in the following Supplementary Text document.

## SUPPLEMENTARY TEXT

### Index

A. Mammalian Data Set	4
B. ERCs on The Original Phylogeny with Short Branches	9
C. Branch Time to Protein-Rate Correlation Problem	10
D. Partial Correlation to Address BT-PR Correlation	12
E. Removing Short Branches to Remove BT-Rate Confounding Effects	13
F. Testing Whether Branch Rate Increases with Evolutionary Time	17
G. Testing for Taxonomic Order Effects	21
H. ACE2 CRR Subnetworks	23
I. Additional ACE2 Interactor Descriptions	24
J. Supplementary Text References	34

### Supplementary Tables

**Table S1:** Index of all published supplementary materials.

**Table S2:** Disambiguated proteins from OrthoDB groups with ancient family expansions.

**Table S3:** List of taxa considered in the mammalian phylogeny with no adjustment, 20MY adjustment, and 30MY adjustment.

**Table S4:** The top 40 ERCs for ACE2 based on the original ERC method, the BTime-adjusted partial ERC method, and the 30MY adjusted ERC method.

**Table S5:** Select Spearman's correlations for protein rates against BTime using the original rate data.

**Table S6:** Regression and ANOVA results for linear models with ACE2 and BTime as independent variables.

**Table S7:** Select Spearman's correlation test results for rate-to-BTime residual models against BTime.

**Table S8:** Select Spearman's correlation test results for rates against BTime using the original rate data, 20MY-adjusted rate data, and 30MY-adjusted rate data.

**Table S9:** Comparison of the FGA-FGB ERC using the original ERC method, the BTime-adjusted partial ERC method, and the 30MY adjusted ERC method.

**Table S10:** Comparison of the FGA-FGG ERC using the original ERC method, the BTime-adjusted partial ERC method, and the 30MY adjusted ERC method.

**Table S11:** Comparison of the FGB-FGG ERC using the original ERC method, the BTime-adjusted partial ERC method, and the 30MY adjusted ERC method.

**Table S12:** Comparison of the IFNAR1-IFNAR2 ERC using the original ERC method, the BTime-adjusted partial ERC method, and the 30MY adjusted ERC method.

**Table S13:** Comparison of the Collagen Type IV interacting pairs using the original ERC method, the BTime-adjusted partial ERC method, and the 30MY adjusted ERC method.

**Table S14:** Select Wilcoxon matched signed-rank test results testing for whether the rates of branches increased as time scales are increased.

**Table S15:** Linear regression fit statistics for select proteins using models that used ACE2 rate ranks, BTime ranks, and taxonomic order as independent variables.

**Table S16:** Select ANCOVA results for linear models that used ACE2 rate ranks, BTime ranks, and taxonomic order as independent variables.

**Table S17:** Spearman independent contrasts tests for select proteins against ACE2.

## Supplementary Figures

**Figure S1:** Full time-scaled mammalian phylogeny.

**Figure S2:** 30MY-adjusted time-scaled mammalian phylogeny.

**Figure S3:** Scatterplots depicting the rates of ACE2 to some proteins of interest using the original rate data.

**Figure S4:** Scatterplots depicting the rates of ACE2 to some proteins of interest using the original rate data and 30MY-adjusted rate data side-by-side.

**Figure S5:** Cartoon illustrating how branches were extended to test for rate changes as time scales were increased.

**Figure S6:** Box plots depicting the rate shifts following 20MY and 30MY adjustments for select proteins.

**Figure S7:** The distribution of p-values of Wilcoxon matched signed-rank tests of all proteins comparing the rate shifts following the 20MY and 30MY adjustments.

## A. Mammalian Data Set

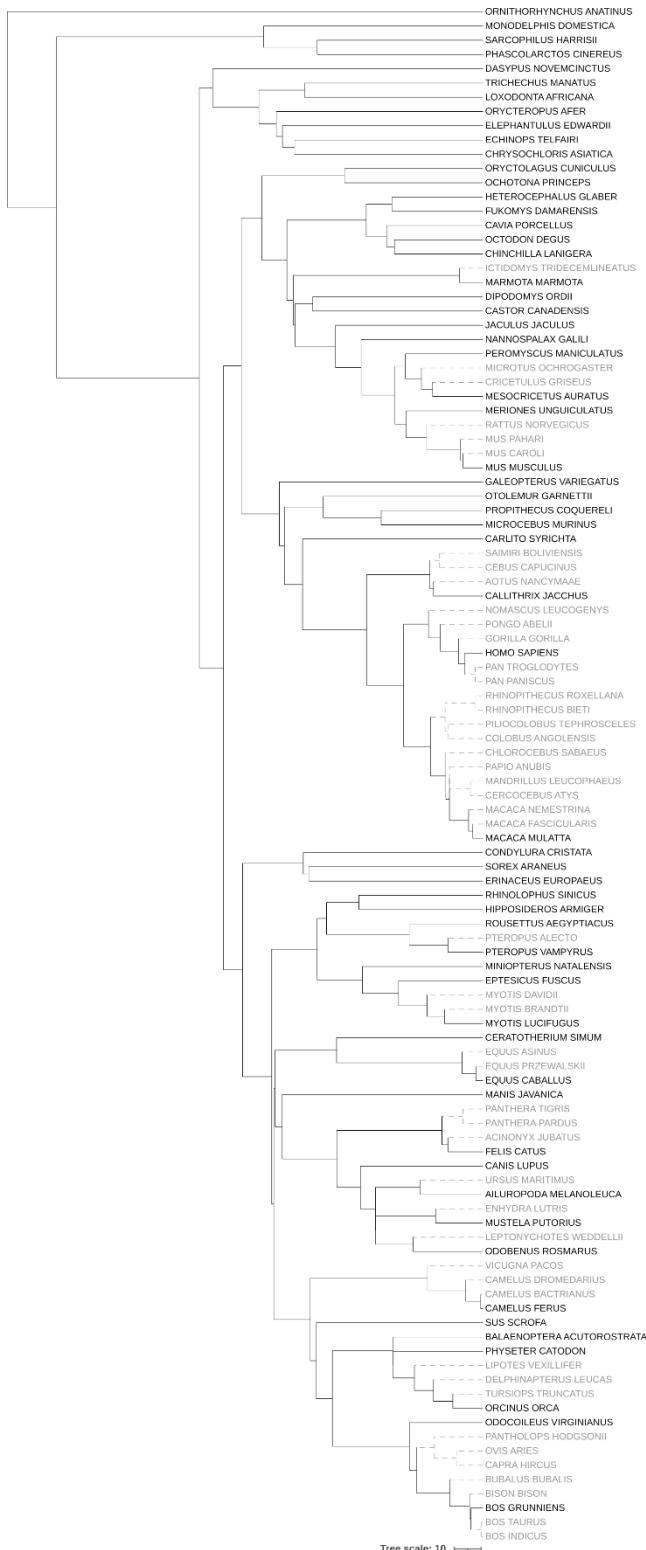
As described in the methods section, the data set is primarily based on the orthologous protein groups available on OrthoDB (Kriventseva et al., 2019) based on the “mammalia” taxonomic level. We selected protein groups that are single-copy in all species with greater than 90 taxa represented. An additional 156 proteins, which did not meet the initial single copy in all taxa requirement, were added to extend the analysis in pathways of interest (e.g. coagulation cascade, sphingolipid signaling, renin-angiotensin system). Of these proteins, 47 were added due to literature suggesting an association with COVID-19, to evaluate their ERCs to ACE2, such as IFNAR2 and XCR1 (Severe Covid-19 GWAS Group, 2020; Pairo-Castineira et al., 2021; Fricke-Galindo & Falfán-Valencia, 2021). Only proteins with relatively minor paralogy issues were added by this method (Supplementary File 5). The rationale for this approach is that it would be very difficult to determine which paralog to choose for the analysis in terminal branches with multiple paralogs for a particular protein. The final set contains a total of 1,953 proteins, including ACE2.

In 23 cases (Table S2), OrthoDB orthology groups contain multiple distinct protein groups resulting from ancient gene duplications. In some cases, we examined the phylogeny of the orthology group and, where appropriate, divided and added them to our protein set. In most cases, the division was supported by protein annotation names within the orthology group, and the protein sequences were split based on reference annotations given by OrthoDB and sequence similarity. For example, coagulation factor IX (F9) and X (F10) were within the same orthology group (OrthoDB ID: 91794at40674).

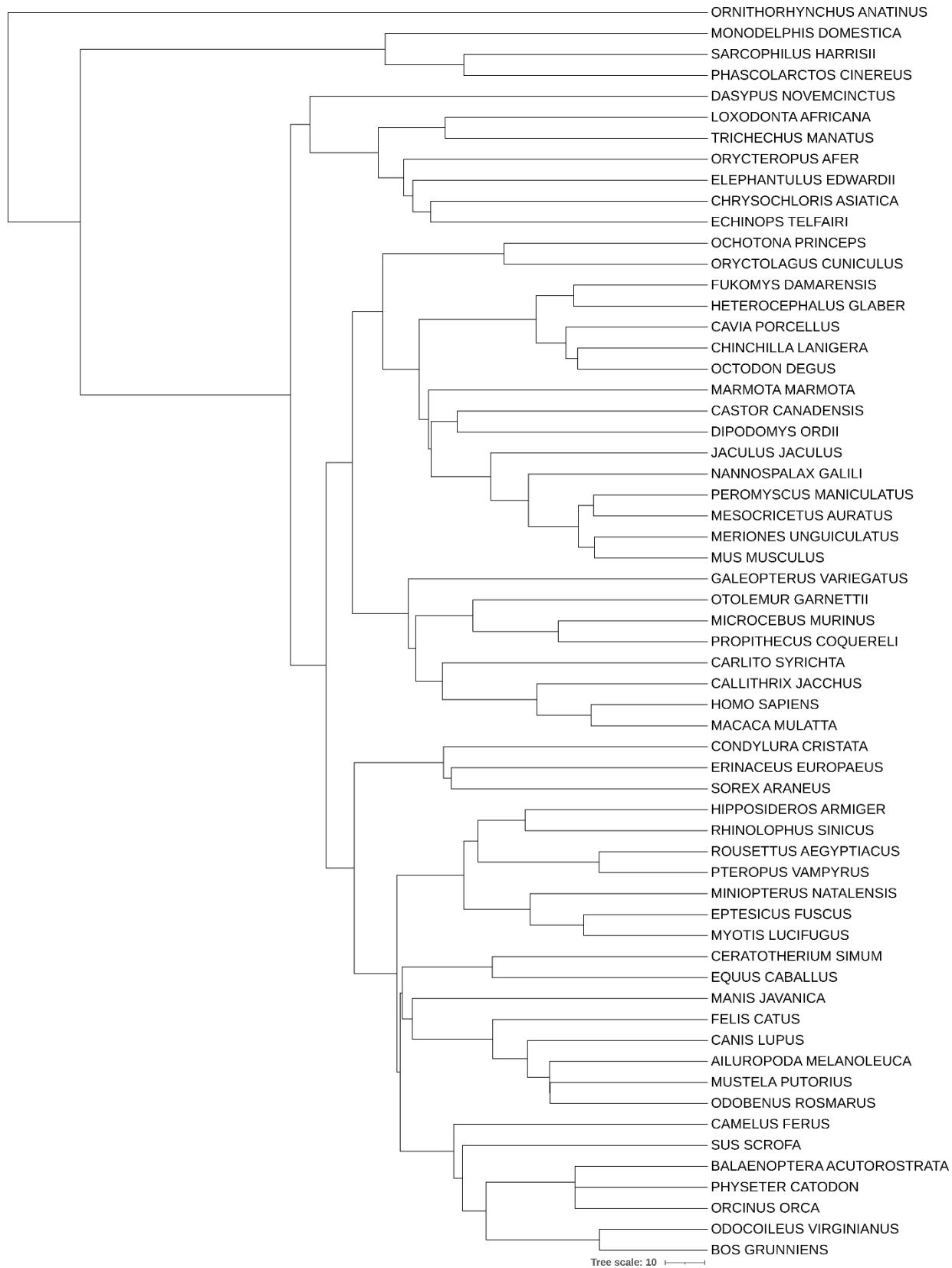
OrthoDB ID	Distinct Proteins Added
10776at40674	IGF1R,INSR
15742at40674	ABCC1,ABCC3,ABCC6
25854at40674	DPP8,DPP9
32671at40674	MAP3K5,MAP3K15
46864at40674	LIFR,OSMR
55743at40674	PRKCI,PRKCZ
66003at40674	BMX,BTK
68344at40674	SPTLC2,SPTLC3
79978at40674	PPP2R5D
85041at40674	TMPRSS2,TMPRSS3
91794at40674	F9,F10
94914at40674	PPP2R2A,PPP2R2B,PPP2R2C
95740at40674	MAPK8,MAPK10
103747at40674	GLA,NAGA
111203at40674	MAPK12,MAPK13
114138at40674	CERS5,CERS6
123408at40674	DEGS1,DEGS2
123688at40674	SGPP1,SGPP2
123726at40674	PPP2CA,PPP2CB
129864at40674	MAPK11,MAPK14
132357at40674	SPHK1,SPHK2
138259at40674	CCR2,CCR5
166274at40674	ACER1,ACER2

**Table S2:** The OrthoDB groups that were added to the dataset for which there were multiple distinct proteins reported as a single orthology group. The proteins listed on the right column were all the disambiguated proteins added to the 30MY dataset (so they had to have met our requirement of having at least 50 of the selected taxa).

A well-resolved time-scaled mammalian phylogeny available from TimeTree (Kumar et al., 2017) was used that includes the taxa that were in our orthologous protein sets. This tree contained 108 mammals (Fig. S1, Table S3) in the original uncorrected data set. Later, in order to correct a terminal branch time (BT) to protein rate correlation found for most proteins due to short branches (see below), we removed taxa from oversampled clades with short terminal branches. We found that a 30MY threshold for terminal branches eliminated the terminal branch time to protein rate for 87.5% of proteins (described in Section E), resulting in 50-60 taxa per protein (Table S3). These data were used for the ERC analysis reported in the main text.



**Figure S1:** Full original phylogeny topology with branches scaled to time (in millions of years) based on TimeTree (Kumar et al., 2017). Branches highlighted in grey are removed following a 30MY branch length threshold correction. The tree illustration is created using iTOL (Letunic & Bork, 2021).



**Figure S2:** Time-scaled phylogeny only containing the 60 selected taxa following a 30MY threshold correction. The tree illustration is created using iTOL (Letunic & Bork, 2021).

Full Taxa List	20MY Taxa List	30MY Taxa List
<i>Ornithorhynchus anatinus</i>	<i>Ornithorhynchus anatinus</i>	<i>Ornithorhynchus anatinus</i>
<i>Dasyurus novemcinctus</i>	<i>Monodelphis domestica</i>	<i>Monodelphis domestica</i>
<i>Orycterus afer</i>	<i>Phascolartos cinereus</i>	<i>Phascolartos cinereus</i>
<i>Elephantulus edwardii</i>	<i>Sarcophilus harriii</i>	<i>Sarcophilus harriii</i>
<i>Echinos telfairi</i>	<i>Dasyurus novemcinctus</i>	<i>Dasyurus novemcinctus</i>
<i>Chrysochloris asiatica</i>	<i>Loxodonta africana</i>	<i>Loxodonta africana</i>
<i>Trichechus manatus</i>	<i>Trichechus manatus</i>	<i>Trichechus manatus</i>
<i>Loxodonta africana</i>	<i>Orycterus afer</i>	<i>Orycterus afer</i>
<i>Galeopterus variegatus</i>	<i>Elephantulus edwardii</i>	<i>Chrysochloris asiatica</i>
<i>Carito syrichta</i>	<i>Chrysochloris asiatica</i>	<i>Echinos telfairi</i>
<i>Nomascus leucogenys</i>	<i>Echinos telfairi</i>	<i>Ornitho princeps</i>
<i>Pongo abelii</i>	<i>Otoodon degus</i>	<i>Otoodon degus</i>
<i>Gorilla gorilla</i>	<i>Oryctolagus cuniculus</i>	<i>Oryctolagus cuniculus</i>
<i>Homo sapiens</i>	<i>Fulmarus glacialis</i>	<i>Fulmarus glacialis</i>
<i>Pan troglodytes</i>	<i>Heterocephalus glaber</i>	<i>Heterocephalus glaber</i>
<i>Pan paniscus</i>	<i>Cavia porcellus</i>	<i>Cavia porcellus</i>
<i>Piliocolobus tephroceles</i>	<i>Chinchilla lanigera</i>	<i>Chinchilla lanigera</i>
<i>Colobus angolensis</i>	<i>Otoodon degus</i>	<i>Otoodon degus</i>
<i>Rhinopithecus roxellana</i>	<i>Marmota marmota</i>	<i>Marmota marmota</i>
<i>Rhinopithecus bieti</i>	<i>Castor canadensis</i>	<i>Castor canadensis</i>
<i>Chirocebus sabaeus</i>	<i>Dipodomys ordii</i>	<i>Dipodomys ordii</i>
<i>Macaca nemestrina</i>	<i>Jaculus jaculus</i>	<i>Jaculus jaculus</i>
<i>Macaca mulatta</i>	<i>Nannopax galili</i>	<i>Nannopax galili</i>
<i>Macaca fascicularis</i>	<i>Peromyscus maniculatus</i>	<i>Peromyscus maniculatus</i>
<i>Papio anubis</i>	<i>Microtus ochrogaster</i>	<i>Mesocricetus auratus</i>
<i>Mandrillus leucophaeus</i>	<i>Cricetulus griseus</i>	<i>Meriones unguiculatus</i>
<i>Cercocebus atys</i>	<i>Mesocricetus auratus</i>	<i>Mus musculus</i>
<i>Saimiri boliviensis</i>	<i>Meriones unguiculatus</i>	<i>Galeopterus variegatus</i>
<i>Cebus capucinus</i>	<i>Rattus norvegicus</i>	<i>Otolemur garnettii</i>
<i>Callithrix jacchus</i>	<i>Mus musculus</i>	<i>Microcebus murinus</i>
<i>Aotus nancymaae</i>	<i>Galeopterus variegatus</i>	<i>Propithecus coquereli</i>
<i>Otolemur garnettii</i>	<i>Otolemur garnettii</i>	<i>Carito syrichta</i>
<i>Propithecus coquereli</i>	<i>Microcebus murinus</i>	<i>Callithrix jacchus</i>
<i>Microcebus murinus</i>	<i>Propithecus coquereli</i>	<i>Homo sapiens</i>
<i>Marmota marmota</i>	<i>Carito syrichta</i>	<i>Macaca mulatta</i>
<i>Ictidomys tridecemlineatus</i>	<i>Aotus nancymaae</i>	<i>Condylura cristata</i>
<i>Dipodomys ordii</i>	<i>Callithrix jacchus</i>	<i>Erinaceus europaeus</i>
<i>Castor canadensis</i>	<i>Saimiri boliviensis</i>	<i>Sorex araneus</i>
<i>Jaculus jaculus</i>	<i>Nomascus leucogenys</i>	<i>Hippopotamus amphibius</i>
<i>Nannopax galili</i>	<i>Homo sapiens</i>	<i>Rhinolophus sinicus</i>
<i>Peromyscus maniculatus</i>	<i>Rhinopithecus bieti</i>	<i>Rousettus aegyptiacus</i>
<i>Microtus ochrogaster</i>	<i>Macaca mulatta</i>	<i>Pteropus vampyrus</i>
<i>Mesocricetus auratus</i>	<i>Condylura cristata</i>	<i>Miniopterus natalensis</i>
<i>Cricetulus griseus</i>	<i>Erinaceus europaeus</i>	<i>Eptesicus fuscus</i>
<i>Meriones unguiculatus</i>	<i>Sorex araneus</i>	<i>Myotis lucifugus</i>
<i>Rattus norvegicus</i>	<i>Hippopotamus amphibius</i>	<i>Ceratotherium simum</i>
<i>Mus pahari</i>	<i>Bulinolophus sinicus</i>	<i>Equus caballus</i>
<i>Mus musculus</i>	<i>Rousettus aegyptiacus</i>	<i>Manis javanica</i>
<i>Mus caroli</i>	<i>Pteropus vampyrus</i>	<i>Felis catus</i>
<i>Cavia porcellus</i>	<i>Miniopterus natalensis</i>	<i>Canis lupus</i>
<i>Otolemur garnettii</i>	<i>Eptesicus fuscus</i>	<i>Mustela erminea</i>
<i>Chinchilla lanigera</i>	<i>Myotis evotis</i>	<i>Mustela putorius</i>
<i>Heterocephalus glaber</i>	<i>Myotis evotis</i>	<i>Odobenus rosmarus</i>
<i>Fulmarus glacialis</i>	<i>Leptonychotes weddellii</i>	<i>Camelus ferus</i>
<i>Oryctolagus cuniculus</i>	<i>Odobenus rosmarus</i>	<i>Sus scrofa</i>
<i>Otolemur princeps</i>	<i>Vicugna pacos</i>	<i>Balaenoptera acutorostrata</i>
<i>Miniopterus natalensis</i>	<i>Camelus ferus</i>	<i>Phylseter catodon</i>
<i>Eptesicus fuscus</i>	<i>Canis lupus</i>	<i>Orcinus orca</i>
<i>Myotis davidi</i>	<i>Ursus maritimus</i>	<i>Odocoileus virginianus</i>
<i>Myotis lucifugus</i>	<i>Enhydra lutris</i>	<i>Bos grunniens</i>
<i>Myotis brandtii</i>	<i>Mustela putorius</i>	
<i>Rousettus aegyptiacus</i>	<i>Leptonychotes weddellii</i>	
<i>Pteropus vampyrus</i>	<i>Odobenus rosmarus</i>	
<i>Pteropus alecto</i>	<i>Delphinapterus leucas</i>	
<i>Rhinolophus sinicus</i>	<i>Orcinus orca</i>	
<i>Hippopotamus amphibius</i>	<i>Odocoileus virginianus</i>	
<i>Sus scrofa</i>	<i>Pantholops hodgsonii</i>	
<i>Odocoileus virginianus</i>	<i>Ovis aries</i>	
<i>Pantholops hodgsonii</i>	<i>Bos grunniens</i>	
<i>Ovis aries</i>		
<i>Capra hircus</i>		
<i>Bubalus bubalis</i>		
<i>Bison bison</i>		
<i>Bos grunniens</i>		
<i>Bos taurus</i>		
<i>Bos indicus</i>		
<i>Balaenoptera acutorostrata</i>		
<i>Phylseter catodon</i>		
<i>Lipotes vexillifer</i>		
<i>Delphinapterus leucas</i>		
<i>Tursiops truncatus</i>		
<i>Orcinus orca</i>		
<i>Vicugna pacos</i>		
<i>Camelus dromedarius</i>		
<i>Camelus ferus</i>		
<i>Camelus bactrianus</i>		
<i>Ceratotherium simum</i>		
<i>Equus asinus</i>		
<i>Equus przewalskii</i>		
<i>Equus caballus</i>		
<i>Manis javanica</i>		
<i>Canis lupus</i>		
<i>Ursus maritimus</i>		
<i>Alurropoda melanoleuca</i>		
<i>Odobenus rosmarus</i>		
<i>Leptonychotes weddellii</i>		
<i>Mustela putorius</i>		
<i>Enhydra lutris</i>		
<i>Felis catus</i>		
<i>Acinonyx jubatus</i>		
<i>Panthera tigris</i>		
<i>Panthera pardus</i>		
<i>Condylura cristata</i>		
<i>Sorex araneus</i>		
<i>Erinaceus europaeus</i>		
<i>Monodelphis domestica</i>		
<i>Sarcophilus harriii</i>		
<i>Phascolartos cinereus</i>		

**Table S3:** List of taxa that are in the original phylogeny (left column), the taxa that are chosen following a 20MY correction (center column), and the taxa which are chosen following the 30MY threshold correction (right column).

The final data set is composed of 1,953 orthologous protein groups with each individual protein containing 50 to 60 taxa total.

## B. ERCs on The Original Phylogeny with Short Branches

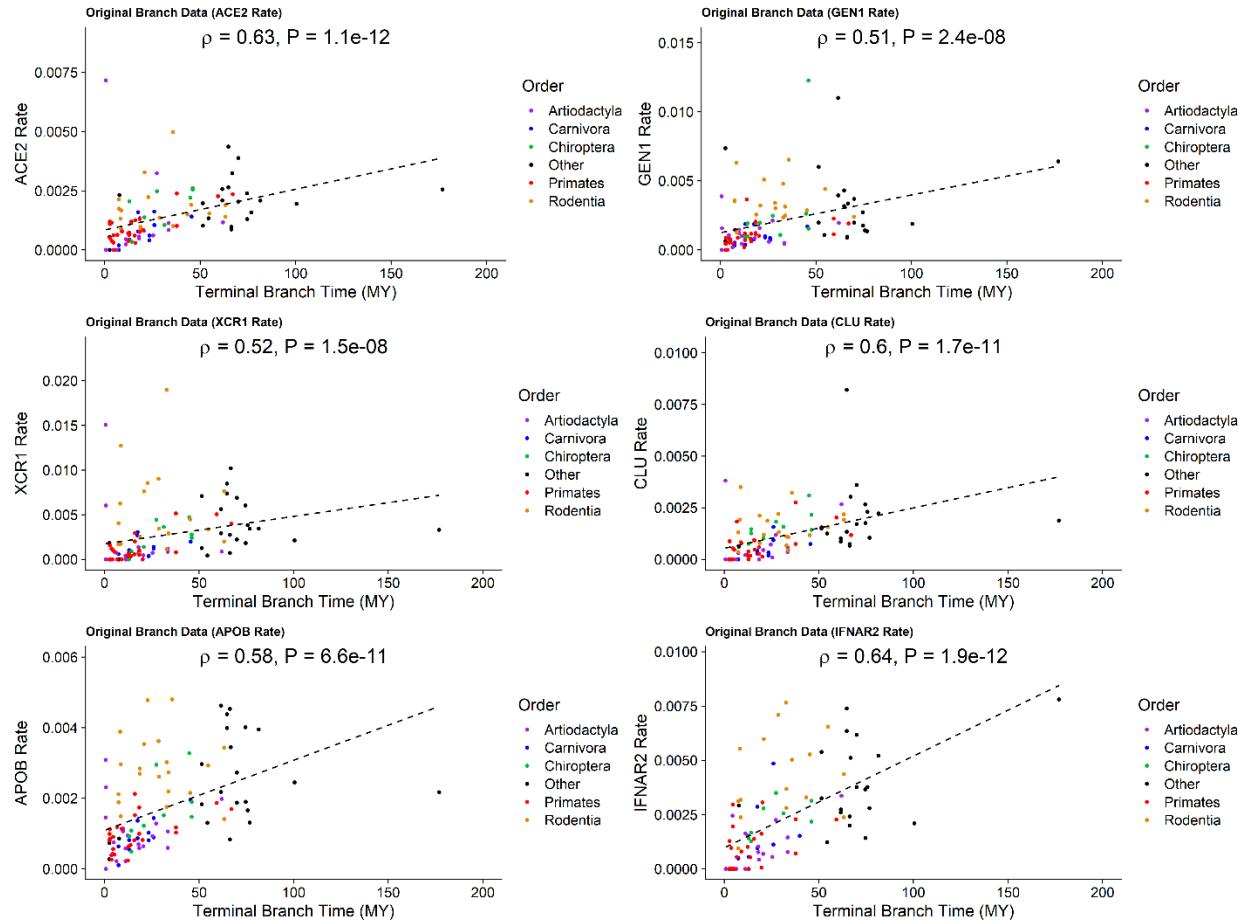
ERCs were initially calculated for the 1,953 proteins using the complete mammalian phylogeny (Fig. S1) using the same scheme as defined in Methods section of the main text. The top 40 ERCs for ACE2 using this initial method are shown in Table S4. However, these ERCs could be driven (in part) by a spurious correlation to branch time (Section C) An initial attempt to remove the correlation was conducted using partial correlations(Kim, 2015) (Section D). The top 40 ACE2 ERCs for this treatment are also presented in Table S4, along with the final, 30MY threshold corrected ERCs. There are 7 proteins (TNFSF18, IFNAR2, GPR141, CLU, F5, SERPINA5, and SLC10A6) that are shared among all three top 40 ACE2 ERCs. Nine proteins are shared between the top 40 original ACE2 ERCs (TSGA13, CLU, F5, GPR141, PLA2G7, SLC10A6, IFNAR2, TNFSF18, and SERPINA5) and the 30MY ERCs, with 8 proteins that are shared between the top 40 ACE2 branch time-corrected ERC and 30MY ERC sets (CLU, F5, COL4A4, GPR141, SLC10A6, IFNAR2, TNFSF18, and SERPINA5).

Rank	Protein	Original ERCs				BT-Corrected ERCs				30MY-Adjusted ERCs				
		$\rho$	P	FDR	P	Protein	$\rho$	P	FDR	P	Protein	$\rho$	P	FDR
1	IFNAR2	0.711	2.0E-15	8.8E-13		PLA2R1	0.567	6.2E-10	1.2E-06		GEN1	0.669	4.3E-08	4.2E-05
2	APOB	0.709	5.5E-17	1.1E-13		APOB	0.543	3.8E-09	3.7E-06		XCR1	0.669	3.2E-08	4.2E-05
3	TNFSF18	0.700	9.9E-16	6.5E-13		CERS3	0.537	6.0E-09	3.9E-06		CLU	0.631	3.1E-07	1.5E-04
4	OSMR	0.699	5.8E-16	5.6E-13		IFNAR2	0.521	1.2E-07	3.3E-05		TMEM63C	0.630	2.0E-07	1.3E-04
5	CERS3	0.682	2.3E-15	8.8E-13		WRN	0.500	8.5E-08	3.3E-05		IFNAR2	0.616	2.5E-06	6.1E-04
6	SERPINA5	0.661	1.3E-13	2.1E-11		RMI1	0.499	1.1E-07	3.3E-05		KIF3B	0.599	1.7E-06	4.9E-04
7	PLA2G7	0.658	7.5E-14	1.8E-11		SPHKAP	0.498	1.0E-07	3.3E-05		ITPRIP1L2	0.590	1.7E-06	4.9E-04
8	LIFR	0.657	1.1E-13	2.1E-11		GPR183	0.495	1.4E-07	3.4E-05		FAM227A	0.589	1.8E-06	4.9E-04
9	SLC51B	0.656	6.9E-13	5.4E-11		OSMR	0.486	3.5E-07	6.0E-05		TLR8	0.583	3.7E-06	7.2E-04
10	GPR183	0.655	8.5E-14	1.8E-11		TNFSF18	0.483	5.5E-07	7.6E-05		COL4A4	0.579	3.7E-06	7.2E-04
11	FGB	0.655	6.5E-14	1.8E-11		COL4A4	0.483	3.7E-07	6.0E-05		FAM3D	0.574	5.8E-06	8.4E-04
12	F5	0.654	7.2E-14	1.8E-11		SLC10A6	0.482	3.3E-07	6.0E-05		F5	0.572	4.1E-06	7.2E-04
13	RMI1	0.651	1.2E-13	2.1E-11		F5	0.480	3.3E-07	6.0E-05		AR	0.572	7.7E-06	8.8E-04
14	CPB2	0.649	3.6E-13	3.9E-11		EPB42	0.479	4.0E-07	6.1E-05		TSGA13	0.569	7.1E-06	8.8E-04
15	CLU	0.649	4.9E-13	4.2E-11		ZNF830	0.478	1.1E-06	1.0E-04		PLA2G7	0.568	6.0E-06	8.4E-04
16	TSGA13	0.649	2.9E-13	3.5E-11		SERPINA5	0.473	1.0E-06	1.0E-04		MMS19	0.564	5.9E-06	8.4E-04
17	PROCR	0.647	1.9E-13	2.9E-11		CIP2A	0.470	6.2E-07	8.1E-05		AMOT	0.562	8.1E-06	8.8E-04
18	ZNF830	0.646	1.6E-12	9.4E-11		C5orf34	0.469	7.6E-07	8.7E-05		L1CAM	0.560	8.6E-06	8.8E-04
19	C5orf34	0.646	2.4E-13	3.2E-11		CAT	0.468	7.1E-07	8.7E-05		PDYN	0.560	7.3E-06	8.8E-04
20	GPR141	0.645	3.2E-13	3.7E-11		MUC15	0.465	1.3E-06	1.1E-04		IQCD	0.559	9.2E-06	8.9E-04
21	CXCL10	0.645	2.5E-13	3.2E-11		VTN	0.464	1.2E-06	1.0E-04		SERPINA5	0.557	2.2E-05	1.4E-03
22	PLA2R1	0.640	4.5E-13	4.2E-11		SELP	0.461	1.1E-06	1.0E-04		CERS4	0.555	2.9E-05	1.5E-03
23	EPB42	0.639	4.8E-13	4.2E-11		PPP1R3A	0.460	1.1E-06	1.0E-04		CC2D1B	0.555	1.1E-05	1.0E-03
24	WRN	0.637	4.7E-13	4.2E-11		PLG	0.459	1.1E-05	3.8E-04		GPR141	0.552	1.5E-05	1.2E-03
25	IFIH1	0.637	4.7E-13	4.2E-11		KITLG	0.455	1.8E-06	1.4E-04		FSCB	0.551	2.8E-05	1.5E-03
26	BDKRB2	0.635	1.3E-12	8.5E-11		FGB	0.454	1.6E-06	1.3E-04		RGR	0.549	3.0E-05	1.5E-03
27	IL1B	0.635	1.3E-12	8.6E-11		FAM237A	0.454	3.0E-06	2.0E-04		COL4A5	0.549	2.1E-05	1.4E-03
28	BVES	0.634	6.3E-13	5.2E-11		GPR141	0.452	2.3E-06	1.7E-04		TNFSF8	0.548	1.2E-05	1.1E-03
29	COL1A2	0.633	1.3E-12	8.5E-11		APOBR	0.451	5.0E-06	2.4E-04		CCDC36	0.548	1.5E-05	1.2E-03
30	CAT	0.632	7.9E-13	6.0E-11		CXCL10	0.450	2.3E-06	1.7E-04		MRC1	0.548	1.3E-05	1.1E-03
31	TLR7	0.632	1.4E-12	8.7E-11		CLU	0.449	4.1E-06	2.1E-04		CD27	0.545	3.0E-05	1.5E-03
32	PPP1R3A	0.632	8.5E-13	6.1E-11		ATP10D	0.448	3.0E-06	2.0E-04		ADCK4	0.545	2.1E-05	1.4E-03
33	HK3	0.629	1.8E-12	1.0E-10		LIFR	0.446	3.7E-06	2.0E-04		SOWAHA	0.543	2.2E-05	1.4E-03
34	SLC10A6	0.629	1.4E-12	8.7E-11		FER1L5	0.446	3.3E-06	2.0E-04		F2RL2	0.539	3.7E-05	1.7E-03
35	VTN	0.625	2.9E-12	1.5E-10		SERTAD4	0.446	2.6E-06	1.8E-04		WDR66	0.536	2.1E-05	1.4E-03
36	BCLAF3	0.624	2.4E-12	1.3E-10		PPP2R3A	0.443	3.6E-06	2.0E-04		TRADD	0.535	2.6E-05	1.5E-03
37	REL	0.623	2.0E-12	1.1E-10		IL5RA	0.442	4.6E-06	2.2E-04		RELA	0.533	2.8E-05	1.5E-03
38	RNASEL	0.621	3.2E-12	1.7E-10		PIGV	0.442	3.7E-06	2.0E-04		SLC10A6	0.531	3.0E-05	1.5E-03
39	APOBR	0.621	2.0E-11	7.1E-10		ZDHHC4	0.442	3.7E-06	2.0E-04		IL23A	0.529	4.7E-05	1.7E-03
40	SELENOP	0.618	5.6E-12	2.7E-10		CPED1	0.441	3.5E-06	2.0E-04		TNFSF18	0.528	5.8E-05	1.8E-03

**Table S4:** The top 40 ERCs for ACE2 based on the original ERC method (left), BT-Corrected partial correlation ERC method (center), and the standard 30MY-adjusted ERC method (right). FDR corrections are based on the full ERC dataset for each respective ERC method.

### C. Branch Time to Protein-Rate Correlation Problem

In examining the terminal branch rate correlation data for ACE2, we found that its rate of evolution was correlated with the terminal branch time (BT) (illustrated in Fig. S3). We suspect that this correlation may be due to episodic selection over the course of its evolution (possibly driven in part by evolution in its partners). As a result, BT could be a confounding correlate in ERC. Examination of the proteins in our set indicated a significant BT correlation to evolutionary rate for 1,559 out of 1,953 proteins ( $p < 0.05$ ; Supplementary File 6). Notably, many of the strongest original ERCs to ACE2 (such as IFNAR2 and APOB), have very significant correlations to BT with  $p$  values greater than 0.5 (Table S5). To directly test the effects of time on predicted ERC interactions, multiple linear regressions were performed on the rank-transformed rate data from protein relationships of interest, with time as a covariate (equations in the form:  $Protein_{RateRank} = \beta_2 ACE2_{RateRank} + \beta_1 BranchTime_{Rank} + \beta_0$ ). Many of the proteins with strong ACE2 ERCs resulted in models with the time variable being a significant factor (Table S6). These results additionally hold using similar models under an ANOVA test (Table S6). Examining scatterplots of protein evolutionary rates indicate that the pattern may be driven by short branches with respect to BT (examples in Fig. S3). As expected by this interpretation, the vast majority of proteins (all but 37 of 1953; Supplementary File 7) show significantly more points below the regression line for short branches (<30MY). The short branches occur in relatively oversampled taxonomic orders, as oversampling of closely related species shortens terminal branch times. Since BT is a significant covariate in the original ERC data, the significant ERCs could be due, in part, to a confounding covariance to BT. We therefore examined different approaches to remove this confounding variable (below).



**Figure S3:** A set of scatterplots depicting the rates of evolution of several proteins of interest plotted against terminal branch time for the original data, with highly sampled clades colored. Also depicted is the linear regression line to emphasize the positive association and Spearman's rank correlation test results ( $\rho$  and  $p$ -value). In each case, the rate data shows a significant correlation with BT. For each protein, there are significantly more points below the regression line for terminal branches  $< 30$  MY, indicating lower rates for short branches.

Protein	$\rho$	P-Value
ACE2	0.629	1.10E-12
GEN1	0.512	2.38E-08
XCR1	0.521	1.45E-08
CLU	0.605	1.71E-11
TMEM63C	0.440	1.87E-06
IFNAR2	0.641	1.92E-12
KIF3B	0.117	2.32E-01
ITPRIPL2	0.534	2.67E-09
FAM227A	0.540	1.57E-09
TLR8	0.459	1.07E-06
COL4A4	0.384	5.23E-05
APOB	0.576	6.64E-11
PLA2R1	0.375	6.87E-05
CAT	0.515	1.20E-08
CERS3	0.498	4.30E-08

**Table S5:** Spearman’s rank correlation tests on the terminal branch rates against BT for branch time uncorrected data to proteins of interest (strong ERCs in the original or 30MY threshold ERCs). In all cases shown, the proteins have a strong correlation between their terminal branch rates and time prior to correction for short branches.

Protein	$R^2_{adj}$	Model P	Intercept Term	Intercept P	ACE2 Term	ACE2 P	Time Term	Time P	ANOVA ACE2 P	ANOVA Time P
GEN1	0.382	2.68E-11	15.672	2.18E-03	0.449	2.48E-05	0.241	1.92E-02	2.00E-11	1.92E-02
XCR1	0.341	6.02E-10	17.340	1.05E-03	0.384	4.37E-04	0.272	1.13E-02	7.58E-10	1.13E-02
CLU	0.476	1.72E-14	11.114	1.66E-02	0.453	4.07E-06	0.323	7.49E-04	5.19E-14	7.49E-04
TMEM63C	0.222	1.33E-06	24.124	4.63E-05	0.204	7.20E-02	0.332	3.94E-03	7.41E-06	3.94E-03
IFNAR2	0.556	7.33E-17	7.939	5.32E-02	0.515	1.18E-07	0.314	6.92E-04	1.33E-16	6.92E-04
KIF3B	0.125	5.21E-04	36.182	2.42E-08	0.458	2.12E-04	-0.168	1.62E-01	2.60E-04	1.62E-01
ITPRIPL2	0.292	1.21E-08	20.694	2.24E-04	0.183	9.12E-02	0.419	1.69E-04	5.37E-07	1.69E-04
FAM227A	0.339	3.94E-10	17.841	9.16E-04	0.330	1.92E-03	0.327	2.14E-03	1.55E-09	2.14E-03
TLR8	0.322	2.93E-09	18.343	6.31E-04	0.433	9.81E-05	0.200	6.37E-02	1.35E-09	6.37E-02
COL4A4	0.327	1.35E-09	21.071	1.04E-04	0.578	3.68E-07	0.008	9.37E-01	1.83E-10	9.37E-01
APOB	0.523	3.15E-17	10.849	1.62E-02	0.569	3.79E-09	0.223	1.29E-02	1.58E-17	1.29E-02
PLA2R1	0.400	3.98E-12	19.693	1.34E-04	0.677	6.17E-10	-0.059	5.49E-01	5.45E-13	5.49E-01
CAT	0.408	1.58E-12	15.654	2.03E-03	0.519	7.13E-07	0.180	6.94E-02	5.22E-13	6.94E-02
CERS3	0.467	8.43E-15	13.804	4.03E-03	0.592	6.02E-09	0.143	1.29E-01	1.79E-15	1.29E-01

**Table S6:** Linear model fit using the original data set to test for branch time and ACE2 effects, using the form:  $Protein_{RateRank} = \beta_2 ACE2_{RateRank} + \beta_1 BranchTime_{Rank} + \beta_0$ . Selected proteins of interest are shown from top ACE2 ERCs of the original and 30MY data sets. In all cases, except for TMEM63C and ITPRIPL2, the model has a strongly significant reported P-value, indicating that ACE2 is significantly predictive. For 8 of 14 proteins branch time is also significantly predictive. For ANOVA, all 14 proteins show a significant ACE2 effect, and 8 of 14 have a significant Branch time effect. This indicates that branch time is a confounding factor for many ACE2’s ERCs in the original data, which contains short terminal branches.

#### D. Partial Correlation to Address BT-PR Correlation

As time is a significant confounding effect on the protein rate, ERCs values may be distorted by the branch time covariate. We, therefore, investigated the use of “partial correlations” to control for the confounding effect of time on our correlation calculations (Kim, 2015). Partial correlation-based ERCs were generated utilizing the “ppcor” R package (Kim, 2015) to produce Spearman’s rank partial correlation tests while controlling for the effects of terminal branch time. The partial correlations are based on fitting a linear model to the variable(s) being controlled for and then

performing a Spearman’s rank correlation test on the residuals of the two models. These residuals represent the variance in the data that are unexplained by the variable(s) being controlled for. In particular, terminal branch time was controlled to account for the observed correlation to BT. Even following the partial correlation controlling for BT, ACE2 still had strong ERCs to immune system-related proteins such as IFNAR2 (Table S4). However, partial correlations are not robust to assumption violations. As partial correlations are based on performing a rank correlation test on the residuals of linear models of rates trained against time, we examined the data to assess the possibility of these violations. Several problems were noted upon examining residuals of individually trained models. The most important of which is that rate vs BT residuals were still correlated with BT. Since these residuals should capture variance that is not explained by terminal branch time, it is unexpected for these residuals to still have a strong association to BT. However, 1,529 of 1,953 proteins have residuals that still have a significant Spearman’s correlation to time ( $p < 0.05$  Supplementary File 8, select proteins are displayed in Table S7). Key proteins such as ACE2 are among the set of proteins with residuals that still correlate significantly to BT (Table S7). The previous analysis showed that short branch rates are overrepresented below the protein rate to branch time regression line for the vast many proteins, which likely explains why the partial regression fails to remove the branch time correlation in many cases.

Protein	Residuals vs Time $\rho$	Residuals vs BTime P
ACE2	0.197	4.60E-02
GEN1	-0.006	9.51E-01
XCR1	0.133	1.77E-01
CLU	0.092	3.55E-01
TMEM63C	0.374	6.55E-05
IFNAR2	0.081	4.35E-01
KIF3B	0.519	1.17E-08
ITPR1PL2	0.365	1.03E-04
FAM227A	0.284	2.86E-03
TLR8	0.200	4.23E-02
COL4A4	0.546	1.75E-09
APOB	0.114	2.41E-01
PLA2R1	0.327	5.75E-04
CAT	0.408	1.14E-05
CERS3	0.299	1.67E-03

**Table S7:** Spearman’s rank correlation tests of the residuals of linear models trained on a protein’s rates against time. Ten of the 15 proteins depicted (including ACE2) retain a significant association with time after accounting for time. Full table available in Supplementary File 8.

As we noted that short branches appear to drive the rate to BT correlation (Fig. S3), we therefore decided to control for confounding branch time effects by removing short branches and recalculating ERC rates.

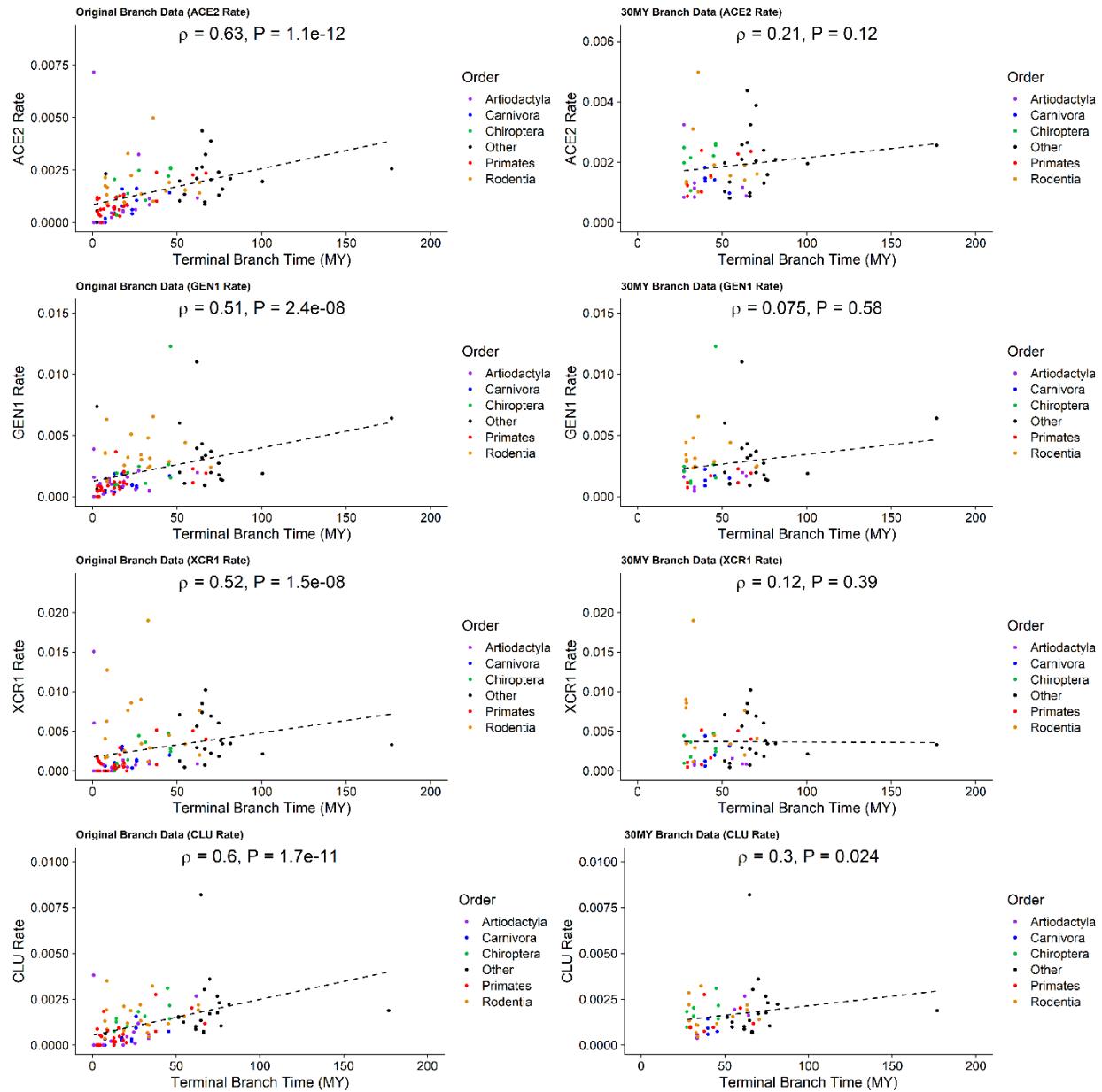
### E. Removing Short Branches to Remove the Confounding BT-Rate Factor

As we observed that terminal branch time is a confounding factor in our ERC analysis (Section C), we examined short branches as a likely driver for the association. Therefore, we identified sister taxa with short branches and selectively remove one or more, to remove short branches and extend branches in the remaining sister taxa (Fig. S2, Table S3). The procedure was applied to produce clades with branch lengths with a 20MY BT threshold or a 30MY BT threshold. Note that we allowed around a 3 MY buffer (e.g. 30-27 MY threshold) so as to not restrict the taxonomic sample sizes too heavily. The specific representative taxa were picked arbitrarily, but generally were chosen to allow for the greatest number of internal nodes to be merged into a single branch

(Table S3), with the main exception to the rule being that *Homo sapiens* was selected as the representative of its clade, due to its relevance to the COVID-19. The taxa selections at the 20MY time scale resulted in the removal of 32 taxa from the original phylogeny and the taxa selections at the 30MY time scale resulted in the removal of 48 taxa. Both adjustments to the data strongly reduced the number of proteins displaying a significant association between rate and BT. Specifically, while the original data set had 1,559 out of 1,953 proteins which displayed a significant correlation between BT and rate ( $p < 0.05$ ), the 20MY adjustment reduced this number to 1,065 proteins, and the 30MY adjustment reduced the number of proteins with a significant rate to BT correlation to 245 (select proteins in Table S8, complete set in Supplementary File 6), or 12.5% of proteins. Therefore, the 30MY terminal branch length threshold most effectively removed branch time as a confounding factor. After the 30MY correction, there is no longer a significant correlation between branch time and branch rate for most proteins, as illustrated in Table S8 and Figure S4.

Protein	Full Rate-Btime		20MY Rate-Btime		30MY Rate-Btime	
	$\rho$	P-Value	$\rho$	P-Value	$\rho$	P-Value
ACE2	0.629	1.10E-12	0.484	1.86E-05	0.208	1.24E-01
GEN1	0.512	2.38E-08	0.381	8.73E-04	0.075	5.78E-01
XCR1	0.521	1.45E-08	0.436	1.32E-04	0.116	3.86E-01
CLU	0.605	1.71E-11	0.553	3.91E-07	0.297	2.36E-02
TMEM63C	0.440	1.87E-06	0.248	3.07E-02	0.189	1.49E-01
IFNAR2	0.641	1.92E-12	0.377	1.94E-03	0.263	5.67E-02
KIF3B	0.117	2.32E-01	0.010	9.35E-01	0.164	2.17E-01
ITPR1PL2	0.534	2.67E-09	0.498	4.60E-06	0.221	8.94E-02
FAM227A	0.540	1.57E-09	0.293	1.03E-02	0.192	1.42E-01
TLR8	0.459	1.07E-06	0.462	4.33E-05	0.241	7.05E-02
COL4A4	0.384	5.23E-05	0.397	5.05E-04	0.132	3.20E-01
APOB	0.576	6.64E-11	0.488	7.69E-06	0.207	1.12E-01
PLA2R1	0.375	6.87E-05	0.309	6.93E-03	0.014	9.14E-01
CAT	0.515	1.20E-08	0.246	3.25E-02	0.046	7.27E-01
CERS3	0.498	4.30E-08	0.316	5.48E-03	-0.048	7.15E-01

**Table S8:** Spearman’s rank correlation tests on the terminal branch rates versus branch time for proteins of interest for the three different time threshold treatments: No cutoff, 20MY cutoff, 30MY cutoff. In all cases, the correlation with rate and time decrease—to the point where unadjusted p-values are insignificant at  $p < 0.05$  level for all but one protein at the 30MY cutoff.



**Figure S4:** A set of scatterplots depicting the rate of evolution of several proteins of interest plotted against terminal branch time with highly sampled clades colored. The left column of plots depicts the original rate data and the right column depicts the corresponding rate data following a 30MY adjustment. Also depicted is the regression line to emphasize the positive association and the statistics of Spearman's rank correlation test results ( $\rho$  and  $p$ -value). In each case, the original data shows a significant correlation with BT while the 30MY adjusted data shows that the association is no longer significant.

Results from the 30MY adjustment also reveal strong reciprocal ERCs among proteins known to occur in complex with each other that were not apparent in the uncorrected ERC analysis. For instance, the three fibrinogen subunits FGA, FGB, and FGG form a well-known fibrinogen complex (Mosesson, 2005), and have strong reciprocal rank ERCs in the 30MY data, but do not

in the original treatment (Tables S9-S11). Similar empirical observations were noted among several other interacting proteins such as the weak relationship between IFNAR1 and IFNAR2 in the uncorrected data but the much stronger relationship in the 30MY data (Table S12), despite their being known to complex (Thomas et al., 2011). We also note weak relationships between several of the Collagen Type IV subunits in the uncorrected ERC data, but the relationships were again strengthened following the 30MY adjustment (Table S13) which are known to physically interact (Casino et al., 2018), and found to form strong reciprocal rank ERCs in the corrected data set.

FGA and FGB				
ERC	$\rho$	Raw P	Rank of FGB for FGA	Rank of FGA for FGB
Original	0.703	8.5E-17	48	6
Time-Corrected	0.575	2.1E-10	62	5
30MY-Corrected	0.696	9.6E-10	12	1

Table S9: The ERC results between the expected interacting proteins FGA and FGB under the original ERC method, the time-corrected partial correlation-based ERC, and the final 30MY-corrected ERC. This interaction does not meet our reciprocal rank 20 criteria until we use the 30MY-corrected ERCs.

FGA and FGG				
ERC	$\rho$	Raw P	Rank of FGG for FGA	Rank of FGA for FGG
Original	0.661	3.8E-14	109	5
Time-Corrected	0.563	9.2E-10	77	5
30MY-Corrected	0.722	1.6E-10	2	1

Table S10: The ERC results between the expected interacting proteins FGA and FGG under the original ERC method, the time-corrected partial correlation-based ERC, and the final 30MY-corrected ERC. This interaction does not meet our reciprocal rank 20 criteria until we use the 30MY-corrected ERCs, additionally, the 30MY ERC value itself is strongest after the 30MY correction.

FGB and FGG				
ERC	$\rho$	Raw P	Rank of FGG for FGB	Rank of FGB for FGG
Original	0.670	5.4E-15	23	4
Time-Corrected	0.568	3.3E-10	11	4
30MY-Corrected	0.603	4.3E-07	4	7

Table S11: The ERC results between the expected interacting proteins FGB and FGG under the original ERC method, the time-corrected partial correlation-based ERC, and the final 30MY-corrected ERC. This interaction does not meet our reciprocal rank 20 criteria using the original ERC calculation. It does meet the reciprocal rank 20 criteria after time correction, but this reciprocal rank interaction gets even stronger after the 30MY correction.

ERC	$\rho$	Raw P	IFNAR1 and IFNAR2	
			Rank of IFNAR2 for IFNAR1	Rank of IFNAR1 for IFNAR2
Original	0.635	1.8E-11	52	215
Time-Corrected	0.497	7.1E-07	131	167
30MY-Corrected	0.786	2.1E-11	2	18

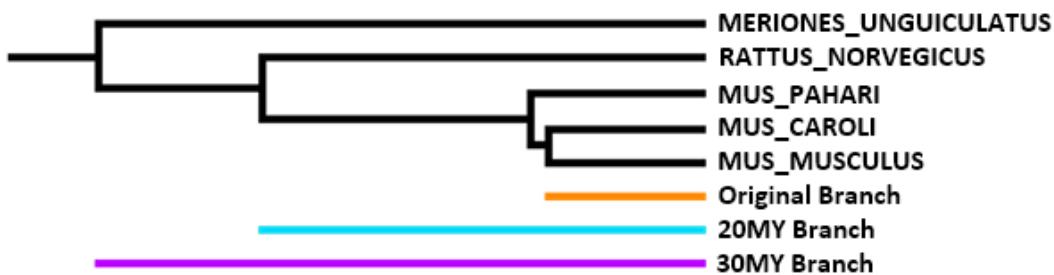
Table S12: The ERC results between the expected interacting proteins IFNAR1 and IFNAR2 under the original ERC method, the time-corrected partial correlation-based ERC, and the final 30MY-corrected ERC. Notably, the interaction does not meet our reciprocal rank 20 criteria until our 30MY correction. We also note that the 30MY ERC is stronger than all other attempts.

Protein Pairs	Interacting Collagen Type IV Pairs		
	Original ERC Ranks	Time-Corrected ERC Ranks	30MY-Corrected ERC Ranks
COL4A1 - COL4A2	134 - 70	157 - 54	7 - 22
COL4A3 - COL4A4	48 - 57	40 - 38	1 - 1
COL4A3 - COL4A5	414 - 262	423 - 303	115 - 84
COL4A4 - COL4A5	228 - 123	189 - 115	105 - 74
COL4A5 - COL4A6	7 - 33	4 - 23	7 - 18

Table S13: The ERC ranks of protein pairs of interacting Collagen Type IV subunits according to Casino et al. (2018) under different ERC corrections. P and p-values are omitted for clarity but in all instances, the  $\rho$  values were increased under the 30MY correction when compared to either the time-corrected or original ERCs.

## F. Testing Whether Branch Rate Increases with Evolutionary Time

There is a positive association between terminal branch time and the rate of evolution for many proteins (Section C). The question, therefore, arises as to whether there is actually an increase in evolutionary rate over time for these proteins. To test this question, we conducted an “experiment” to extend branches along independent clades, in order to test whether increasing branch time increases protein evolutionary rate. This was accomplished by extending branch lengths along taxonomic branches in different clades by trimming adjacent taxa and comparing the protein rates as branches are extended. (Fig. S5). Based on the TimeTree phylogeny (Kumar et al., 2017), we selected individual clades containing short branches that would have their time scales extended following a 20MY and 30MY adjustment (Fig. S5, Table S3).



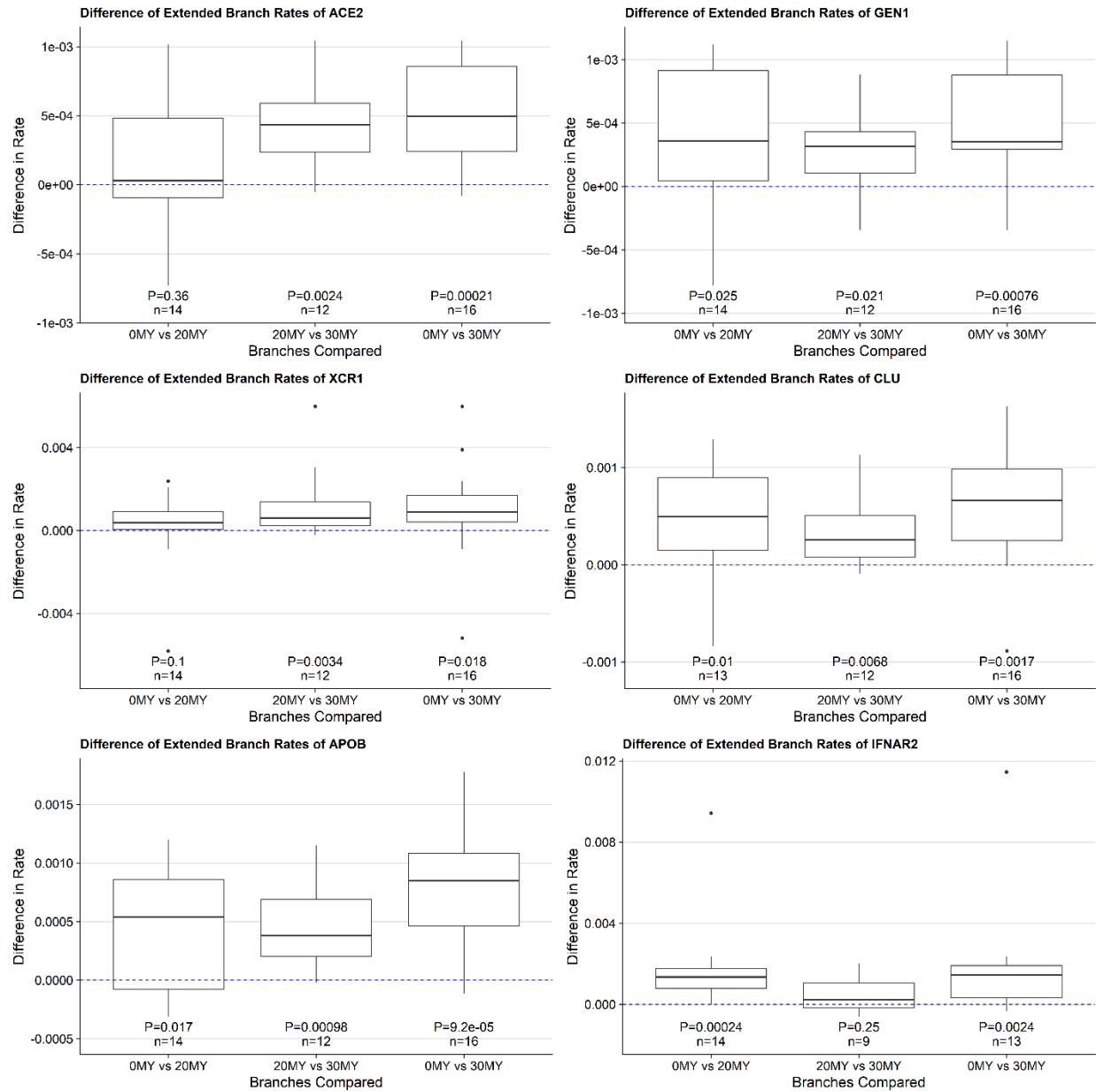
**Figure S5:** Cartoon illustrating the branches being compared when testing whether branch rates change upon an increase in time scale. In this instance, the taxon “Mus musculus” is selected from

the *Rattus* and *Mus* clades. The original short branch (orange), 20MY branch (cyan), and 30MY branch (purple) are each used to calculate rates, and these are the paired data that is compared to test for changes in rates.

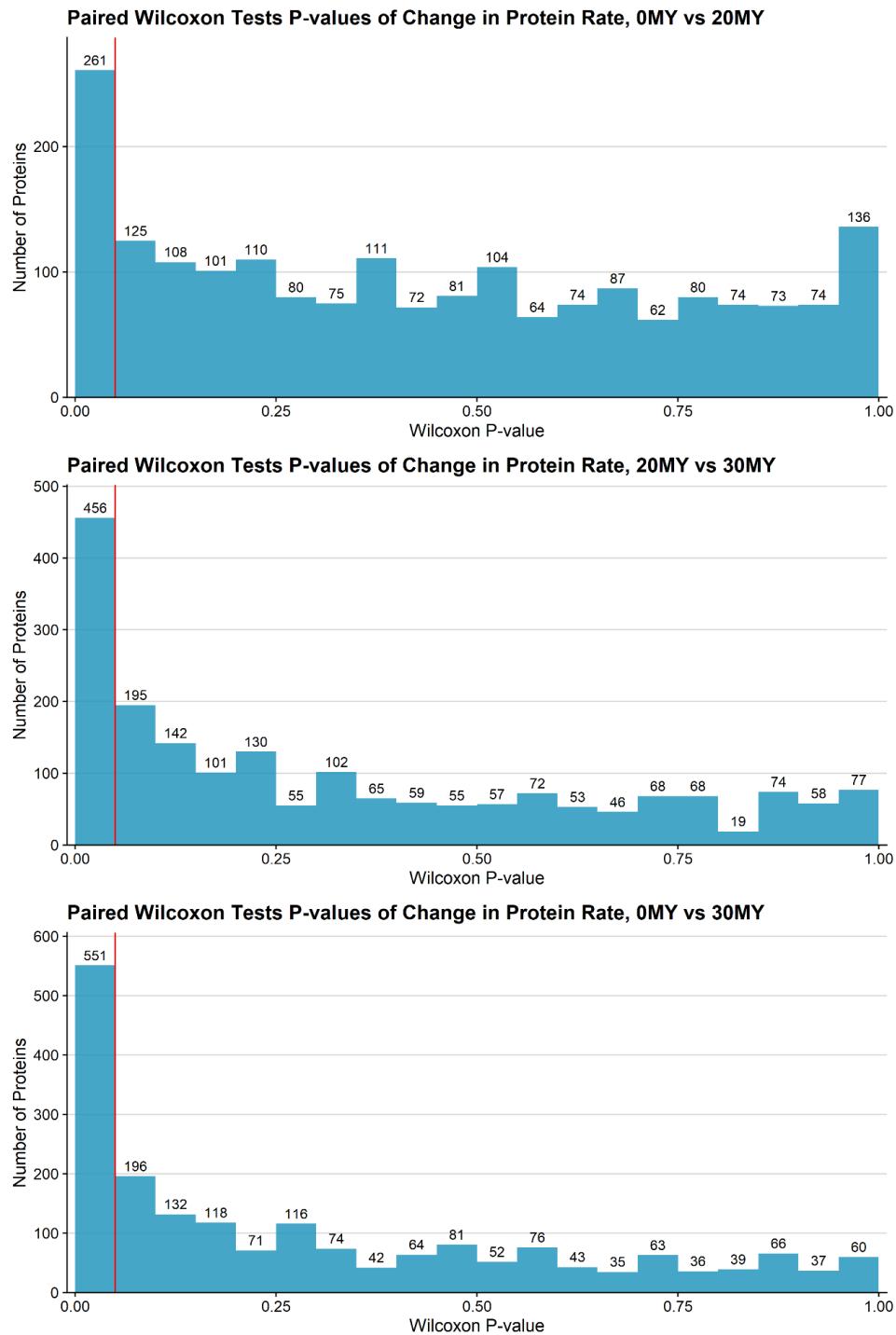
Since we suspected that rates scale as time increases, we specifically tested whether there is a significant difference in rate for each of these branches before and after 20MY and 30MY adjustments, as described in Section E (14 selected taxa for comparing original vs 20MY, 12 selected taxa for comparing 20MY vs 30MY, 16 selected taxa for comparing original vs 30MY). Tests on each branch's rate against the respective adjusted rate were performed using two-tailed Wilcoxon Matched Signed Rank Tests (results for all proteins are reported in Supplementary File 9), to test whether these rates significantly differed. We note that many proteins show significant changes in rate under each adjustment, but this pattern is most prominent in the shift from short branch rates to 30MY rates (longer branches). Examples are shown in Table S14 and Figure S6, and the complete data are present in Supplementary File 9. Notably, out of our set of 1,953 proteins using a significance cutoff of  $p < 0.05$ , 261 proteins show significant rate changes (238 of which have a median increase in rate) in the Short-to-20MY treatment, 456 show significant rate changes (442 of which have a median increase in rate) in the 20MY-to-30MY treatment, and 551 show significant rate changes (545 of which have a median increase in rate) in the Short-to-30MY treatment (Fig. S7).

Protein	Short vs 20MY P	20MY vs 30MY P	Short vs 30MY P
ACE2	3.58E-01	2.44E-03	2.14E-04
GEN1	2.45E-02	2.10E-02	7.63E-04
XCR1	1.04E-01	3.42E-03	1.82E-02
CLU	1.05E-02	6.84E-03	1.68E-03
TMEM63C	4.63E-01	2.33E-01	1.93E-01
IFNAR2	2.44E-04	2.50E-01	2.44E-03
KIF3B	9.52E-01	7.91E-01	8.60E-01
ITPRIPL2	1.00E+00	5.22E-02	4.64E-01
FAM227A	9.52E-01	6.77E-01	9.00E-01
TLR8	3.53E-02	9.77E-04	4.27E-03
COL4A4	5.83E-01	3.22E-02	7.39E-02
APOB	1.66E-02	9.77E-04	9.16E-05
PLA2R1	5.83E-01	9.77E-04	5.77E-02
CAT	8.54E-03	1.10E-01	1.68E-03
CERS3	1.94E-01	4.88E-04	9.19E-03

**Table S14:** Unadjusted P-values for two-tailed paired Wilcoxon signed-rank tests comparing the rate of evolution of selected branches after various adjustments for selected proteins of interest. Most proteins show significant differences in rate, and all but PLA2R1 has a significant difference in rates from the original rate data and 30MY rate data.



**Figure S6:** Boxplots of the differences in the rate of evolution of selected branches after various adjustments for selected proteins of interest. A dashed blue line indicates a difference of zero. Sample size and two-tailed paired Wilcoxon signed-rank test p-values are indicated underneath each respective box indicating if there was a significant change in rates.



**Figure S7:** The distributions of p-values of the two-tailed Wilcoxon matched signed-rank tests comparing whether there is a significant difference in the rates of difference in selected branches when time scales were increased. Additionally, the vertical red line indicates a  $p < 0.05$  threshold for significance, such that all bins to the right of it represent insignificant tests.

We hypothesize that these shifts in rate may be due to increased evolutionary time scales being able to capture episodic evolutionary events that would otherwise be missed in the short branches of the original phylogeny. As longer time scales are considered, there could be a larger chance that these episodic events would be captured, explaining the pattern.

## G. Testing for Taxonomic Order Effects

We use three methods to test for taxonomic order effects on the calculated 30MY ERCs, (1) multiple linear regression, (2) analysis of covariance (ANCOVA), and (3) non-parametric independent contrasts. For the regression and ANCOVA approaches, 30MY rate data is grouped by mammalian taxonomic orders accessed via ETE3 (Huerta-Cepas, Serra & Bork, 2016) and treated as an independent variable. The independent contrasts test uses the mammalian topology previously created with TimeTree (Kumar et al., 2017) to generate independent contrasts within the phylogeny. Statistical tests for each method are performed using base R (version 3.6.1).

Linear regression models using mammalian order as a variable were tested in the following general equation format:  $Protein_{RateRank} = \beta_3 ACE2_{RateRank} + \beta_2 BranchTime_{RateRank} + \beta_1 Order + \beta_0$  on the 30MY adjusted terminal branch time data. Since taxonomic order is a categorical variable, R implicitly converts the variable to become a one-hot encoded “contrast” matrix. One can then examine the reported model metrics to see if any of the encoded taxonomic order variables have a statistically significant contribution to the resultant model. We focus our analysis on the top 5 proteins showing high 30MY ERCs with ACE2 (GEN1, XCR1, CLU, TMEM63C, and IFNAR2). All the examined models have a strong fit (Table S15). In most cases, none of the orders provide a significant contribution to the model (Supplementary File 10). There are a few notable exceptions. The model for GEN1 displays a near-significant contribution of Rodentia, but removing Rodentia still results in a significant ERC to ACE2 ( $p = 0.60$ , unadjusted  $p = 4.5E-11$ ) so the ERC is not an artifact of the effect of Rodentia. Additionally, the model for CLU displays a significant contribution of Dasyuromorphia (Supplementary File 10), however, there is only one taxon within the order in the data and there is still a strong ERC when this taxon is removed ( $p = 0.67$ , unadjusted  $p = 3.2E-08$ ). So, we do not consider this an important contributor to the ACE2-CLU relationship, and it is more likely to be due to model overfitting. We also note that IFNAR2’s model shows a significant contribution of the Carnivora, Cingulata, Perissodactyla, Pholidota, and Primates. (Supplementary File 9). But the ERC between ACE2 and IFNAR2 is still strong after removing these orders from the 30MY rate data ( $p = 0.56$ , unadjusted  $p = 3.7E-04$ ). Importantly, all the models calculated show a significant contribution of ACE2, even in the presence of these order effects ( $p$ -values range from  $2.04E-02$  to  $4.1E-04$ ; Supplementary File 9). Furthermore, the linear models for each of these proteins of interest show an insignificant contribution of branch time using the 30MY-based rate data, further validating the removal of the rate-time correlation (Supplementary File 10).

Protein	Model	
	$R^2_{Adj}$	P-Value
GEN1	0.678	1.04E-06
XCR1	0.520	2.93E-04
CLU	0.416	2.63E-03
TMEM63C	0.456	9.68E-04
IFNAR2	0.724	1.51E-06

**Table S15:** The adjusted  $R^2$  and overall model significance values for each of the linear models representing ACE2’s top 5 ERCs to test for the effects of taxonomic order. In all cases, the model is significant at  $p < 0.05$  and has strong fits reported by the  $R^2$  values, confirming the relationships identified with the 30MY ERCs between ACE2 and these proteins.

As an alternate method to test for the effects of taxonomic order, we used ANCOVA. ANCOVA is a parametric test that allows for the inclusion of categorical data. Since ANCOVA has a similar model structure as linear modeling, the same model structure described above is once again utilized for statistical testing. ACE2’s top 5 ERC partners in the 30MY set have no significant effect of taxonomic order except for GEN1 ( $p = 1.6E-03$ ; Table S16) and IFNAR2 ( $p = 2.5E-04$ ; Table S16). However, ACE2 has a much more significant contribution to each of these models than does Order ( $p = 7.9E-10$  for GEN1 and  $p = 7.8E-09$  for IFNAR2; Table S16). Removing the orders identified above in the regression analysis eliminates the significant order effect detected by ANCOVA for GEN1 ( $p = 2.2E-01$ ) and reduces the effect for IFNAR2 ( $p = 8.6E-03$ ). But as discussed above, the ERCs for ACE2 to GEN1 and to IFNAR2 are still strong and significant after removing the taxa identified in the regression analysis. We also note again, that under 30MY adjustment, terminal branch time is not a significant covariate in all cases examined (Table S16).

Protein	ACE2 P	Order P	BTime P
GEN1	7.9E-10	1.6E-03	7.0E-01
XCR1	4.8E-08	1.3E-01	9.9E-01
CLU	8.2E-07	4.9E-01	6.4E-02
TMEM63C	2.9E-07	1.8E-01	5.8E-01
IFNAR2	7.8E-09	2.5E-04	3.5E-01

**Table S16:** Table showing the p-values for the covariates of ANCOVA tests run on linear models considering the rates of proteins of interest against ACE2 with taxonomic order and terminal branch time.

A Spearman non-parametric independent contrasts test (Garland, Harvey & Ives, 1992) was also used to check for taxonomic effects in the 30MY adjusted rate data. The independent contrasts test is used to examine if there is a significant relationship between ACE2 rates and its top 5 ERC partners even after accounting for taxonomic effects between related species. The test is performed using the R packages “ape” (Paradis & Schliep, 2019) and “picante” (Kembel et al., 2010). In all cases, ACE2 continues to have a significant relationship to each protein ( $p < 0.05$ ), indicating that ACE2’s 30MY ERC relationships are not driven by taxonomic bias (Table S17).

Protein	ACE2 30MY ERCs		ACE2 Contrasts Correlations	
	$\rho$	P-Value	$\rho$	P-Value
GEN1	0.669	4.27E-08	0.673	3.39E-08
XCR1	0.669	3.16E-08	0.503	1.05E-04
CLU	0.631	3.07E-07	0.604	1.32E-06
TMEM63C	0.630	2.02E-07	0.674	1.27E-08
IFNAR2	0.616	2.48E-06	0.613	2.80E-06

**Table S17:** Table showing the correlation coefficients and p-values for the Spearman non-parametric independent contrasts tests on ACE2 against the top 5 ACE2 ERC proteins controlling for phylogenetic effects with the use of independent contrasts. In all cases, the proteins retain a strongly significant correlation with ACE2.

## H. ACE2 CRR Subnetworks

Here we briefly describe other subnetworks within the CRR network with implications to COVID-19.

*TMEM63C RR Subnetwork:* TMEM63C is one of four proteins that form a direct reciprocal rank ERC association with ACE2 (RR20). It functions in osmolarity regulation. In addition to ACE2, TMEM63C has direct RR20 connections to three proteins, CCDC105, LECT2, and C16orf78, and through them forms a subnetwork also containing TMCO2, ARMC7, PAX4, ETS1, and SUV39H1 (Fig. 3). LECT2 (Leukocyte Cell-derived Chemotaxin 2) is involved in macrophage activation, insulin resistance and diabetes, and neutrophil chemotaxis (Yamagoe et al., 1996; Zhang et al., 2018; Takata et al., 2021). TMEM63C and LECT2 are significantly correlated ( $\rho = 0.64$ , FDR = 7.5E-04) with high reciprocal ranks (1,6). Thus, LECT2 is connected to ACE2 through reciprocal ranks between TMEM63C and ACE2.

Little is known about C16orf78, except that it is enriched in testes, and specifically in spermatids (Uhlén et al., 2015). We, therefore, looked at its ERC protein associations as an exploratory tool for possible function. C16orf78 forms a strong RR association with TMEM63C (RR 1,4) and also has reciprocal ERCs with PAX4 (3,9), ETS1 (11,2), and SUV39H1 (10,15). ETS1 is a transcription factor involved in cytokine and chemokine processes. Whereas SUV39H1 is a suppressor of variegation protein whose function loss leads to chromosome instability. It has only 24 significant ERC proteins. These show enrichment for pri-miRNA transcription from RNA polymerase II promoter (FDR = 7.8E-03), scavenging by class A receptors Homo sapiens (FDR = 2.6E-02), endosomal part (FDR = 2.8E-02) and striated muscle tissue development (FDR = 4.9E-02). Pri-miRNAs are processed into miRNAs, whereas scavenging class A receptors play a role in innate immunity as phagocytic receptors in macrophages and dendritic cells (Areschoug & Gordon, 2009), which ties to the endosome term enrichment. These observations may serve as a guide for further investigations into C16orf78 function.

There is also little information on CCDC105, except an intriguing paper using phylogenetic profiling to implicate it as functioning in meiosis-specific chromatin and spermatogenesis (Tabach et al., 2013) and an association of a human variant with infertility (Handel & Schimenti, 2010). Consistent with those two studies, enrichment of the top 2% ERCs for CCDC105 has one

significant term, from proteomicsDB for spermatozoon (FDR = 4.5E-02). It forms strong RR ERCs with the transcription factor PAX4 (3,9) and TMEM63C (1,4).

These findings suggest that TMEM63C protein connections involve innate immunity and spermatogenesis and indicate possible avenues for elucidating interactions of its protein partners of relatively unknown function, such as C16orf78 and CCDC105.

*TSHZ3 Subnetwork:* TSHZ3 does not have a significant ERC to ACE2, yet it connects to ACE2 through FAM3D, with which it has significant reciprocal ranks (1,9). TSHZ3 is a key regulator of airflow and respiratory rhythm control (Caubit et al., 2010), phenotypes that could be important in COVID-19 respiratory distress. Therefore, potential signaling interactions between TSHZ3 and FAM3D3, possibly mediated by physical binding, warrant further examination. Additionally, TSHZ3 variants are associated with amyloid- $\beta$  processing and Alzheimer's disease (Louwersheimer et al., 2017), and it plays a role in smooth muscle development (Caubit et al., 2008). TSHZ3 is highly connected within the ACE2 CRR network, with 10 reciprocal rank connections (Main Text Fig. 3). One of these, BRINP3, connects back to coagulation through FGA and CPB2.

*L1CAM RR Subnetwork:* L1CAM is the fourth protein with a direct RR20 connection to ACE2 (Figure 2). It was originally discovered as an important protein in nervous system development (Moos et al., 1988; Samatov, Wicklein & Tonevitsky, 2016). Among its other functions may be stem cell differentiation to vascular endothelial cells (Rizvanov et al., 2008), and it also plays a role in tumor vascular development (Angiolini & Cavallaro, 2017). These functions may play a role in its protein coevolution with ACE2. Interestingly, BMX is an RR20 to L1CAM and is known as a tyrosine kinase that is present in endothelial and bone marrow cells and may play a role in inflammatory response (Chen et al., 2014). The BMX and ACE2 proteins are encoded by neighboring genes on the X chromosome (Navarro Gonzalez et al., 2021) and BMX has been shown to potentially have two SNPs with strong linkage disequilibrium to an ACE2 SNP associated with the lowered circulation of angiotensin (1-7) (Chen et al., 2018). L1CAM's top 2% ERC enrichment analysis shows significant terms for the coagulation pathway-related (FDR = 5.4E-04), tumor necrosis factor signaling (FDR = 6.4E-03), and gamma-carboxylation of proteins (FDR = 6.7E-03). It connects to FGA through CDKN2C and to FGB through Pfn4.

## I. Additional Information on ACE2 Interactor Proteins

Here we provide additional summary information on ACE2 ERC proteins of interest, based on our review of data sources Gene Cards (Stelzer et al., 2016), KEGG (Kanehisa & Goto, 2000), UniProt (Bateman et al., 2021), NCBI Entrez (Maglott et al., 2005), Human Protein Atlas (Thul et al., 2017), and surveys of literature detected through Google Scholar searches. Additional information on the ERC associations of these proteins is also presented below.

*GEN1 (Flap endonuclease GEN homolog 1):* GEN1 is ACE2's top-ranked ERC ( $p = 0.67$ , FDR = 4.2E-05). It is a DNA nuclease whose primary functions are the resolution of DNA Holliday junctions (Chan & West, 2015), and DNA damage checkpoint signaling (Palmer & Kaldis, 2020). It also has a role in centromere stability in both meiosis and mitosis (Gao et al., 2012). Consistent with its roles in meiosis and mitosis, the second-highest ERC interactor for GEN1 is CC2D1B, a

protein involved in resealing of the nuclear envelope following mitosis and assembly and disassembly of the mitotic spindle (Vietri & Stenmark, 2018).

Surprisingly, the top ERC interactor of GEN1 is Interferon  $\lambda$  receptor 1 (IFNLR1), and they are each other's top-ranked ERC connections (Supplementary File 3). This implies a tight association of GEN1 with the interferon pathways involved in immune response and antiviral defense (Prokunina-Olsson et al., 2020), although there is little evidence for this in the literature. Interferon pathways are important in antiviral defense, but also can contribute to cytokine storms and COVID-19 pathologies (McKechnie & Blish, 2020). Along with SLC10A6 and TESPA1, GEN1, IFNLR1, and CC2D1B form a strong reciprocal rank network (Section D, Figure 3). GEN1's top 2% ERCs are enriched for multiple terms related to viral infection, such as HPV infection (FDR = 2.0E-03), Measles (FDR = 4.0E-03), Hepatitis C (FDR = 4.6E-03), Necroptosis (FDR = 4.6E-03), Influenza A (FDR = 4.7E-03), and Kaposi sarcoma-associated herpesvirus infection (FDR = 5.5E-03). Cytokine-cytokine receptor interaction is another significantly enriched term (FDR = 1.6E-04). In contrast, based on our standard top 2% ERC list for enrichment, there are no significant terms strictly related to DNA replication, despite that being the primary identified function of GEN1 in the scientific literature. We speculate that GEN1's functions in DNA and centrosomes during mitosis could be related to DNA checkpoint signaling affecting apoptosis or necrotic cell death, perhaps explaining the enrichment for proteins involved in viral responses. Identification of binding domains between GEN1 and some of its top ERC partners could be informative for possible functional studies.

*XCR1 (X-C Motif Chemokine Receptor 1):* XCR1 is the 2<sup>nd</sup> top-ranked ERC for ACE2 ( $\rho = 0.67$ , FDR = 6.18E-05). XCR1 is the receptor for the chemokine XCL1. The receptor-cytokine interplay is involved in the immune response to infection and inflammation, development of regulatory T cells in the thymus, and establishing self-tolerance (Lei & Takahama, 2012). Therefore, disruptions of XCR1 due to protein interactions with ACE2 could play a role in COVID-19 complications. As well as being the top rank ACE2 ERC, these two proteins have reciprocal rank correlations at the 2% level (ACE2 is rank 37 for XCR1). Strikingly, the Severe Covid-19 GWAS Group (2020) detected a small genomic region containing six genes that significantly associates with severe COVID-19, one of which is XCR1. Our finding that ACE2's 2<sup>nd</sup> highest ERC interactor is also XCR1 is striking for two reasons. First, it lends independent support for a relationship between COVID-19 and XCR1. Second, it implicates that a direct interaction between ACE2 and XCR1 could be involved in COVID-19 pathologies. To our knowledge, there are no other reports of interactions between these two proteins. Its Top 2% ERCs show an extremely strong enrichment for cytokine-cytokine receptor interactions (FDR = 8.0E-06) and JAK-STAT related terms (FDR = 9.7E-03), and for coagulation and complement and cascades (FDR = 1.0E-02).

*CLU (Clusterin, aka Apolipoprotein J):* CLU is the 3<sup>rd</sup> highest ACE2 ERC ( $\rho = 0.63$ , FDR = 1.5E-04), and these two proteins show strong reciprocal ranks (3, 8), likely supporting biological interactions. Relevant to this point is that both ACE2 and CLU have soluble forms that circulate in the blood (Itakura et al., 2020). CLU prevents aggregation of misfolded proteins in blood by binding to them, and also clears misfolded extracellular proteins by binding to heparan sulfate receptors on cells, leading to endocytosis and degradation of CLU and associated proteins in

lysosomes (Itakura et al., 2020). This recently discovered mechanism has been referred to as a “cleaning squad” for extracellular misfolded proteins (Sánchez-Martín & Komatsu, 2020). CLU also protects cells from complement-induced apoptosis and lysis (Jenne & Tschopp, 1989). As well as being abundant in blood plasma, CLU is also found on mature sperm and abundant in seminal plasma (Uhlén et al., 2015).

CLU shows the strongest possible reciprocal ranking with GPR141 (1,1 -  $\rho$  = 0.68, FDR = 9.1E-06). GPR141 is associated with megakaryocytes (see below). Consistent with their strong evolutionary correlation, CLU is produced in megakaryocytes which subsequently mature into platelets (Tschopp et al., 1993). CLU is released by activated platelets in surrounding fluids at sites of vascular injury (Witte et al., 1993), which is consistent with their function in reducing protein aggregations. A surprising finding is the significant association of Clusterin with several coagulation pathway-related proteins (ranks shown in parentheses), including: F5 (3), F13B (9), FGG (18), and FGA (27). In addition, it has a strong reciprocal interaction with mitochondrial malic enzyme 2 (ME2,  $\rho$  = 0.62, FDR = 3.9E-05, reciprocal ranks 12,2). Analysis of CLU’s top 2% strongest ERCs shows significant enrichment for 186 terms. CLU’s top 4 most significantly enriched terms all relate to the coagulation cascades and clot formation. Additional significant terms are relevant to immunity, such as “Immune system” (FDR = 4.8E-03), “Signaling by Interleukins” (FDR = 4.1E-03), and “Plasma Cell”, an activated immune cell type (FDR = 3.4E-05).

Of direct relevance to COVID-19, Singh et al (2021) found in an expression study of coronavirus infected cells that SARS-CoV-2, SARS-CoV, and MERS-CoV, show shared expression alterations for two genes, one of which is CLU. Therefore, the ERC results for CLU are consistent with aspects of their known function, and their interactions with coronavirus infections.

*GPR141 (G Protein-Coupled Receptor 141):* Although GPR141 falls just outside the top 1% ACE2 ERC set (rank 24 – 1.2%), its relevance to Clusterin and our protein network analysis below warrants its inclusion here. There is limited information on GPR141 in the literature. Nevertheless, GPR141 forms a very strong reciprocal rank with CLU (1,1), each being the top interactor with the other, and CLU-GPR141-ACE2 forms a reciprocal rank 24 triad. According to the Human Protein Atlas (Uhlén et al., 2015), it is highly expressed in the brain, bone marrow, lymphatic tissue, and blood. Cell types showing enriched expression of GPR141 include granulocytes, Kupffer cells, and macrophages, as well as alveolar cell types 1 & 2. A recent study found that GPR141 expression is a molecular signature for megakaryocytes (Lu et al., 2018), the progenitor cells for platelets and red blood cells. Noteworthy in this regard is that autopsy results of COVID-19 victims with neurological manifestations find an unusual presence of megakaryocytes in brain capillaries (Nauen et al., 2021). Additionally, elevated levels of IFN-activated megakaryocytes are observed in the blood of patients with severe COVID-19 (Bernardes et al., 2020). These findings suggest possible roles for GPR141 in COVID-19 pathologies.

Although there is limited information on GPR141, its protein interactions revealed by ERCs could be informative. The GPR141’s top 2 percent ERCs show significant enrichment for 111 terms (Supplementary File 3). Most of its top enriched terms relate to the coagulation cascade (FDR = 2.9E-10), with many of the contributing proteins being similar to Clusterin’s protein set.

Additionally, there is significant enrichment for terms related to regulation of vasodilator nitric oxide (FDR = 3.0E-03), ceramide/sphingolipid signaling (FDR = 6.8E-03) and cytokine responses (FDR = 6.8E-03).

Recent studies implicate GPR141 in Alzheimer's Disease (AD) (Srinivasan et al., 2020; Hodges et al., 2021; Novikova et al., 2021). The finding may be noteworthy given the very strong ERC association of GPR141 with CLU and their top reciprocal ranks (1,1). Multiple lines of evidence implicate CLU in AD, including a role in amyloid A $\beta$  processing, CLU polymorphism association with late-onset AD (Balcar et al., 2021), and correlations of CLU levels in serum and cerebrospinal fluid with AD (Shepherd et al., 2020). Since the function of GPR141 is poorly understood, the ERC results suggest that the two proteins interact closely, possibly through physical binding, and their functional relationships should be further explored.

*TMEM63C (Transmembrane Protein 63C):* TMEM63C is the 4<sup>th</sup> ranking ACE2 ERC (FDR = 1.3E-04), and the two have strong reciprocal ranks (and ACE2 show a strong reciprocal rank ERCs (3,10), suggestive of direct reciprocal interactions. Along with other family members, TMEM63C forms a membrane channel and functions in osmolarity perception and regulation (Zhao et al., 2016). It plays an important role in kidney function and kidney disease (Schulz et al., 2019), with angiotensin II inducing its expression in glomerular podocyte cells (Eisenreich et al., 2020). Reduced expression of TMEM63C can result in podocyte apoptosis (Eisenreich et al., 2020). The connection between TMEM63C and angiotensin II is a further indication of a functional interaction, given that ACE2 metabolizes angiotensin II to angiotensin (1-7) as part of the RAS pathway. The RAS pathway is implicated in aspects of COVID-19 (Kai & Kai, 2020).

TMEM63C's top 2% ERC list has significant enrichment for three terms related to the coagulation cascade (FDR = 6.8E-04). Tissue enrichment reveals "adult liver" as the most enriched term (FDR = 8.0E-03). Importantly, there are significant terms related to peptidase activity and the Renin-angiotensin system (driven by the proteins ACE2 and ANPEP). ANPEP is particularly interesting as it has been previously identified as a receptor for several coronaviruses such as HCV-229E (Yeager et al., 1992). ANPEP is known to be a metallopeptidase (as is ACE2) and has been implicated in the regulation of angiogenesis (Rangel et al., 2007). Additionally, ANPEP is known to have Angiotensin III as a substrate (Danziger, 2008), tying it back to the RAS pathway, with ACE2 and TMEM63C. Therefore, the ACE2-TMEM63C reciprocal rank ERCs may indicate direct biological interactions between the proteins, possibly involving physical binding.

*IFNAR2 (Interferon alpha/beta receptor 2):* IFNAR2 is the 5<sup>th</sup> ranking ACE2 ERC, with highly significant correlation ( $\rho = 0.62$ , FDR = 6.1E-04). IFNAR2 combines with IFNAR1 to form the IFN-alpha/beta receptor, which acts through JAK/STAT signaling to modulate immune responses. IFNAR1/IFNAR2 is the receptor for both alpha and beta interferons and is involved in immune responses to viral infection, most notably to influenza and defense against bacterial infections (Shepardson et al., 2018). IFNAR2 was not originally in our protein set, but we added it based on a paper that implicated this protein in severe COVID-19 based on GWAS and gene expression changes (Liu et al., 2021; Pairo-Castineira et al., 2021). Another study implicates mutations in IFNAR2 with severe COVID-19 (Zhang et al., 2020). When added to our ERC protein set, it was

found to be a high ERC to ACE2 (rank 5 in the ACE2 set), providing independent support for its role in COVID-19, possibly through direct ACE2-IFNAR2 interactions.

There are both soluble and membrane-bound forms of IFNAR2. The soluble form (sIFNAR2) “exerts immunomodulatory, antiproliferative and antiviral activities” (Hurtado-Guerrero et al., 2020). The presence of soluble forms for both IFNAR2 and ACE2 suggests possible avenues for physical interaction, in addition to between their membrane-bound forms. IFNAR2 and IFNAR1 combine to form the IFN-alpha/beta receptor, and as expected, these two proteins are significantly and highly correlated ( $\rho = 0.79$ , FDR = 1.9E-09, reciprocal ranks 19,2). CD40, which ranks IFNAR2 as its top ERC, is a crucial immunity protein in the tumor necrosis factor-R (TNF-R) family, with roles in B lymphocytes, macrophages, and cytotoxic T lymphocytes (Grewal & Flavell, 1998; Van Kooten & Banchereau, 2000). IFNAR2 has eleven proteins showing RR20, which is discussed further in the analysis of reciprocal rank networks (Section D). Enrichment analysis for IFNAR2’s top 2% ERCs has an expected strong enrichment for terms related to canonical IFNAR2-related pathways such as “Cytokine-cytokine receptor interaction” (FDR = 1.4E-04), “PI3K-Akt Signaling pathway” (FDR = 1.8E-03), and “JAK-STAT signaling pathway” (FDR = 4.0E-03). Some additional enriched terms of note include several terms related to: tumor necrosis factor signaling, coagulation and complement cascade, ECM receptor interaction/collagen function, and plasma membrane (Supplementary File 3).

*KIF3B* (*Kinesin Family Member 3B*): KIF3B is the 6<sup>th</sup> highest ACE2 ERC. This protein is involved in chromosomal segregation during meiosis and mitosis and also participates in intracellular trafficking (Stelzer et al., 2016). Along with GEN1, it is another high-ranking ACE2 ERC involved in chromosomal processes. Among its phenotypes are ciliary assembly (Cogné et al., 2020), endocytosis (Reed et al., 2010), and regulation of dendrite structure in neurons (Joseph et al., 2020). KIF3B’s top ERC is Secretogranin II (SCG2), which is a neuroendocrine protein that regulates the formation of secretory granules (Stelzer et al., 2016). Genetic variants of its 2<sup>nd</sup> ranking ERC, Inositol hexakisphosphate kinase 3 (IP6K3) are associated with Alzheimer’s disease (Crocco et al., 2016) and its 4<sup>th</sup> ranking protein, Neuronal Pentraxin Receptor (NPTXR), with which it has strong reciprocal ranks (4,6), is a biomarker for Alzheimer’s disease (Lim et al., 2020). The nature of KIF3Bs interactions with ACE2 is not immediately obvious, except for a possible functional connection between ACE2 at amyloid protein catalysis (Kehoe, 2018; Evans et al., 2020). KIF3B top 2% ERCs show significant enrichment only for the “Complement and coagulation cascades” term from KEGG (FDR = 1.9E-02).

*ITPRIPL2* (*Inositol 1,4,5-Trisphosphate Receptor Interacting Protein-Like 2*): ITPRIPL2 is the 7<sup>th</sup> highest among ACE2’s ERC set. Information about this protein is limited in the literature. It is reported in the Human Protein Atlas to be localized to centrosomes. Examination of its ERC set could provide some information relevant to studies of this protein and possible interactions with ACE2. Among its highest ranking ERCs are two proteins associated with DNA repair and mitotic processes. FANCG (1) is involved with double-strand break repair (Yamamoto et al., 2003). CC2D1B plays a role in the reformation of the mitotic nuclear envelope (Vietri & Stenmark, 2018), has a high reciprocal rank association with ITPRIPL2 (2,6). In turn, CCD1B has high reciprocal ranks with GEN1 (2,1), which is involved in holiday junction resolution and genomic stability (see description above). These findings are consistent with the centrosome localization of ITPRIPL2

and suggest that these proteins may physically interact in a manner that results in correlated protein evolution. Three other proteins showing reciprocal rank associations (RR10) are CC2D1B (2, 6), ENAM (4,5), and STAT6 (10,9). Why ACE2 shows a high ERC with ITPRIPL2 is unclear. An ITPRIPL2 top 2% ERC enrichment analysis indicates cytokine receptor activity (FDR = 1.6E-02) and tumor necrosis factor signaling terms (FDR = 2.4E-02). Additionally, there is significant enrichment for “DNA metabolic process” (FDR = 4.9E-02).

*FAM227A (Family with Sequence Similarity 227 Member A):* FAM227A is the 8<sup>th</sup> ranking ACE2 ERC. There is little information about this protein in the current literature, so its evolutionary protein correlations could be informative. The Human Protein Atlas indicates that gene expression is enhanced in the pituitary gland and testes, in ciliated cells, early and late spermatids, and cone & rod photoreceptors (Uhlén et al., 2015). The top five ERC proteins for FAM227A are F5 (involved in blood coagulation), SPZ1 (enriched in spermatids), C16orf96 (enriched in spermatids), FSCB (enriched in spermatids ), and FERIL5 (enriched in spermatids) (Uhlén et al., 2015). This ERC pattern strongly suggests functional interactions among these proteins in spermatogenesis. Moreover, ACE2 is expressed in spermatogonia (Wang & Xu, 2020) and is implicated in male fertility issues associated with COVID-19 (Liu et al., 2020; Verma, Saksena & Sadri-Ardekani, 2020). Therefore, we suggest that this effect could be mediated by FAM227A, a possibility that is worth further exploration. The top 2% of FAM227A ERCs are enriched for 40 terms and reveal a strong association with inflammatory signaling/immunity (Supplementary File 3). In particular, the most significant enrichment is the KEGG term “Cytokine-cytokine receptor interaction” (FDR = 1.3E-04). Most of the proteins driving enrichment for such terms are toll-like receptors, interferon/interleukin receptors, and cytokine receptors.

*TLR8 (Toll-like Receptor 8):* TLR8 is the 9<sup>th</sup> ranking ACE2 ERC. Toll-like receptors are a class of proteins that can detect and initiate an innate immune response to foreign invaders (Takeda, Kaisho & Akira, 2003) by recognizing conserved features of pathogens (Kawai & Akira, 2010). Importantly, toll-like receptor responses are usually associated with large inflammatory responses of the immune system (Takeda, Kaisho & Akira, 2003; Kawai & Akira, 2010). TLR8 has strong ERCs to several other toll-like receptors such as TLR9 (ranks 11, 13) and a unidirectional connection to TLR7 (rank 26,  $\rho = 0.71$ , FDR = 9.6E-08). Consistent with these observations, enrichment of the top 2% ERC list of TLR8 shows highly significant terms associated with TLR8 such as TRAF6 mediated IRF7 activation in TLR7/8 or 9 signaling (FDR = 8.3E-07) and the toll-like receptor signaling pathway (FDR = 2.1E-06). Additionally, the other significantly enriched terms are overwhelmingly related to other immunity-related pathways (Supplementary File 3).

*COL4A4 (Collagen Type IV Alpha 4):* COL4A4 is the 10<sup>th</sup> ranking ACE2 ERC. Collagen Type 4 is a complex of six proteins that are part of the extracellular matrix called the basement membrane, which resides between epithelial cells (Stelzer et al., 2016), such as those of glomerulus and capillaries. Type 4 collagen is a major constituent of glomerular basement membranes. Mutations in COL4A4 and other COL4A genes are associated with inherited kidney disease such as Alport syndrome (Buzzo et al., 2001) and familial hematuria (Longo et al., 2002). Top 2% ERC list enrichment analysis shows significant enrichment for immunity signaling related terms such as Cytokine-cytokine receptor interaction (FDR = 1.7E-04), PI3k-Akt signaling pathway (FDR =

3.0E-03; of which type IV collagen subunits are canonically annotated as a part of), and JAK-STAT signaling pathway (FDR  $p = 7.0\text{E-}03$ ).

*FAM3D (FAM3 Metabolism Regulating Signaling Molecule D):* FAM3D is the 11<sup>th</sup> ranking ACE2 ERC. As seen in figure ACE2-RRN Net, FAM3D is one of four proteins with strong reciprocal rank correlations to ACE2. It is a chemoattractant for neutrophils and monocytes in peripheral blood, is implicated in inflammatory responses in the gastrointestinal tract (Peng et al., 2016). Studies indicate that it has a role in nutritional regulation in the gastrointestinal tract (de Wit et al., 2012), and this may provide a functional connection, given the role of ACE2 in the processing of peptides in the gut (Kuba et al., 2010). Strikingly, ACE2 and FAM3D show strong ERC reciprocal ranks and form a RR network with CLU and GPR141. It also shows strong RR with Solute Carrier Family 16 Member 11 (SLC16A11). Several coagulation cascade proteins are present in its top 1% interaction set, including F13B (its highest-ranked ERC), SERPINA5, and FGB, suggesting possible links to coagulation pathologies of COVID-19. The top 2%ERC list enrichment analysis results in the top 5 terms related to coagulation and clotting (FDR = 3.5E-09). Additionally, there is strong enrichment for various immune response-related terms such as “cytokine receptor activity” (FDR = 2.2E-03) and enrichment for plasma cell presence (FDR = 5.0E-03).

*F5 (Coagulation Factor 5, also abbreviated FV):* F5 is the 12<sup>th</sup> ranking ACE2 ERC. F5 is a key regulator of hemostasis and a central cofactor involved in blood coagulation (Ivanciu et al., 2017). Our ERC analysis predicts strong interactions between ACE2 and F5 (rank 12 for ACE2,  $p = 0.57$ , FDR = 7.2E-04), possibly mediated through the Clusterin (see below). F5 can act as a cofactor for coagulation or anticoagulation (Cramer & Gale, 2012). Approximately 20% of circulating F5 resides in platelets with the remainder in plasma (Gould, Silveira & Tracy, 2004), and whereas plasma F5 has an important role in thrombin formation in microcirculation, platelet F5 has a larger role in severe injury (Ivanciu et al., 2017). The former role could be relevant to micro thrombosis observed in COVID-19. In fact, F5 has been found to associate with COVID-19 symptom severity (elevation in F5 activity) and this may be due to the high abundance of megakaryocytes in the lungs and hearts in COVID-19 infected patients (Stefely et al., 2020). This is further supported by a gene set overlap study showing F5 being annotated to all five examined comorbidities linked to COVID-19 severity (Dolan et al., 2020).

Our ERC analysis of F5 suggests that it may have many other functions beyond the coagulation pathway. F5 is a very “connected” protein with strikingly strong ERC correlations. Twenty-one proteins have spearman rank correlations  $> 0.80$ . In addition, seven proteins rank F5 first among their ERCs and 43 rank F5 in their top 5 ERCs. The strongest enrichments of the top 2% ERCs are immune response-related terms such as “response to cytokine” (FDR = 1.1E-03) and “inflammatory response” (FDR = 1.2E-03). Notably, there is only one significant coagulation-related term in this list, “Complement and Coagulation Cascades” (FDR = 6.9E-03)

*AR (Androgen Receptor):* AR is the 13th ranked ACE2 ERC ( $p = 0.52$ , FDR = 8.8E-04) and is barely cut off from the RR20 criteria to ACE2 (the rank of ACE2 is 22nd in the AR ERC list). AR is encoded on the X chromosomes and is a hormonal receptor that plays a major role in male development, particularly in male reproductive systems and somatic differentiation (Matsumoto et al., 2008). Its AR’s top-ranking ERC is spermatogenesis associated 25 protein (SPATA25) with

(1,2) reciprocal ranks, and its top 2% ERCs only show significant enrichment for cytokine receptor activity (FDR = 1.1E-03). In addition to its roles in sexual differentiation and behavior (Cunningham, Lumia & McGinnis, 2012), AR enhances prostate cancer cell growth (Gelmann, 2002). It may play a role in microbial infection resistance as a knockout in mice can reduce the development and proliferation of neutrophils (Chuang et al., 2009). Androgen signaling may play a role in SARS-CoV-2 infectivity, as indicated by knockdowns of AR in prostate cells result in downregulation of ACE2 and infection cofactors TMPRSS2 and FURIN (Samuel et al., 2020). Additionally, AR has been annotated as being associated with 4 of the 5 COVID comorbidities that are associated with COVID severity in Dolan et al (2020). Male fertility problems may be associated with COVID-19 infection and the ACE2 receptor is abundant in male genetical track and spermatagonia (Huang et al., 2021; Seymen, 2021). ACE2-AR protein interactions, as predicted by ERC, may play a role in these pathologies.

*TSGA13 (Testis specific gene 13 protein):* TSGA13 is the 14<sup>th</sup> ranking ERC for ACE2 ( $\rho = 0.57$ , FDR = 8.8E-04). The function of this protein is not well understood, so it is characterized by its expression in the testes (Zhao et al., 2015). Despite its high expression in the testes, TSGA13 is expressed in other tissues (Zhao et al., 2015) and it may not play a role in fertility as mice with TSGA13 knocked out were still fertile (Miyata et al., 2016). However, this protein is highly conserved (Zhao et al., 2015) so may still play an important role in organisms. TSGA13 variation has been associated with total colonic aganglionosis in patients with Hirschsprung disease (Jung et al., 2019) and reduced expression of TSGA13 has been associated with carcinoma (Zhao et al., 2015). We, therefore, propose that ERC analysis can provide insight into the potential function of TSGA13 as it has many extraordinarily high ERCs (78 proteins show  $\rho$  values of 0.7 or higher). The top ERC is C16orf96 ( $\rho = 0.83$ , FDR = 4.5E-12) which is not well understood, but its 2<sup>nd</sup> highest ERC is C3orf30 ( $\rho = 0.82$ , FDR = 2E-10), also known as “testis expressed 55” (TEX55) which may play a role in fertility, especially considering its strong expression in adult testes (Jamin et al., 2021). The ERC results coupled with known expression profiles suggest that TSGA13 and C3orf30 may interact with each other, although there is no external evidence to suggest this currently. Furthermore, TSGA13’s potential interaction with ACE2 may be mediated through their common ERC partners such as F5 ( $\rho = 0.80$ , FDR = 3.1E-11), TLR8 ( $\rho = 0.78$ , FDR = 5.7E-10), and IFNAR2 ( $\rho = 0.75$ , FDR = 6.1E-09). The top 2% ERCs show enrichment for many immunity/interferon-related terms (FDR = 7.0E-05), complement and coagulation cascade (FDR = 1.4E-04), and no terms related to male fertility or male reproductive tissues.

*PLA2G7 (Platelet-activating factor acetylhydrolase):* PLA2G7 is the 15<sup>th</sup> ranking ERC for ACE2 ( $\rho = 0.57$ , FDR = 8.4E-04). PLA2G7 is a member of the arachidonic acid pathway and is potentially associated with prostate cancer (Vainio et al., 2011). PLA2G7’s strong ERC to ACE2 is particularly interesting due to its likely association with cardiovascular and heart disease (Sutton et al., 2008; Wang et al., 2010), each of which are associated with COVID-19 (Bansal, 2020; Alsaeid et al., 2020). Additionally, PLA2G7’s role in the arachidonic acid pathway is relevant to COVID-19 pathologies as a deficiency in arachidonic acid may lead to greater COVID-19 susceptibility and the arachidonic acid pathway is a candidate therapeutic target (Hoxha, 2020; Ripon et al., 2021). The connection to ACE2 specifically may also make biological sense as MAS (the receptor for the Angiotensin(1-7) that ACE2 can produce) can cause the release of arachidonic

acid (Bader, 2013). Analysis of PLA2G7's top 2% ERC list shows significant enrichment for various terms related to immunity such as “cytokine receptor activity” (FDR = 1.9E-05) and several viral infection pathways such as Influenza A infection (FDR = 6.5E-03). Interestingly, there was also significant enrichment for terms related to DNA repair (FDR = 4.3E-02).

*MMS19 (MMS19 nucleotide excision repair homolog)*: MMS19 is the 16<sup>th</sup> ranking ERC for ACE2 ( $\rho = 0.56$ , FDR = 8.9E-04). Like ACE2's strongest ERC partner, GEN1, MMS19 is involved in DNA repair (Stehling et al., 2012). It is also specifically associated with the “cytosolic Fe-S protein assembly (CIA)”, which forms a complex with MMS19 to assist in DNA metabolism, replication, and repair (Gari et al., 2012). Similar to GEN1, MMS19's mode of interaction with ACE2 is still unclear. But the top ERCs of MMS19 show several proteins directly related to DNA maintenance such as POLL (DNA polymerase lambda;  $\rho = 0.76$ , FDR = 7.2E-10) and GEN1 ( $\rho = 0.74$ , FDR = 6.2E-09). But significant enrichment on the top 2% ERC list is just shown for “death receptor activity” (FDR = 3.1E-02) and “tumor necrosis factor-activated receptor activity” (FDR = 3.1E-02).

*Angiomotin (AMOT)*: AMOT is the 17<sup>th</sup> ranking ERC for ACE2 ( $\rho = 0.56$ , FDR = 8.8E-04). Its potential relevance to COVID-19 pathologies is clear as AMOT is associated with angiogenesis and endothelial cell movement (Bratt et al., 2005; Aase et al., 2007). These associations may explain its ERC to ACE2 as well. For instance, ACE2 can promote endothelial cell migration (Jin et al., 2015). Additionally, COVID-19 infection has been associated with angiogenesis in the lungs (Ackermann et al., 2020). AMOT shares several of ACE2's top ERCs. For instance, GEN1 and TSGA13 are both among AMOT's top 20 ERCs. The top 2% ERCs of AMOT show significant enrichment for complement and coagulation cascades (FDR = 4.3E-04), inflammatory response (FDR = 1.0E-03), and spermatogenesis (FDR = 2.7E-02).

*L1CAM (L1 cell adhesion molecule)*: L1CAM is a RR20 protein to ACE2 ( $\rho = 0.56$ , FDR = 8.8E-04, ranks 18, 14). It is a part of the immunoglobulin superfamily and is best characterized for its role in the nervous system, specifically relating to the development of the brain (Schäfer & Altevogt, 2010). Interestingly, L1CAM is embedded in the extracellular membrane but can be cleaved near the membrane to allow for the circulation of the truncated protein (Schäfer & Altevogt, 2010). The metallopeptidase ADAM17 is one of the enzymes that cleaves L1CAM near the membrane (Schäfer & Altevogt, 2010), and is also known to mediate the release of the ectodomain of ACE2 from the extracellular membrane as well (Lambert et al., 2005). Thus, both proteins circulate in plasma where they may interact, although the functional basis of this postulated interaction is unclear. L1CAM has three other RR20 proteins: BMX non-receptor tyrosine kinase (BMX; ranks 1,3), cyclin-dependent kinase inhibitor 2C (CDKN2C ranks 2,20), and glycerophosphodiester phosphodiesterase domain containing 3 (GDPD3, 5,19). The top 2% enrichment for L1CAM has several significant terms for complement and coagulation cascades (FDR = 5.4E-04), positive regulation of cellular protein localization (FDR = 5.2E-03), endopeptidase activity (FDR = 5.9E-03), Alzheimer's Disease (FDR = 1.1E-02), and arachnoid cyst (FDR = 3.6E-04). It is possible, although highly speculative, that ACE2-L1CAM protein interactions could play a role in neurological pathologies associated with COVID-19.

*PDYN (Prodynorphin aka Leumorphin)*: PDYN is the 19<sup>th</sup> ranking ERC for ACE2 ( $\rho = 0.56$ , FDR = 8.8E-04). PDYN is an endogenous opioid receptor (Stelzer et al., 2016), which also inhibits vasopressin secretion (Yamada et al., 1988), suggesting a connection to ACE2 in blood pressure homeostasis. Unsurprisingly, PDYN is implicated in neurotransmission and mental disorders (such as schizophrenia, Alzheimer's, epilepsy, and cerebellar ataxia) (Clarke et al., 2012; Jezierska et al., 2013; Henriksson et al., 2014). PDYN has several proteins involved in immune function among its top ERCs such as Interferon lambda receptor 1 (IFNLR1;  $\rho = 0.77$ , FDR = 6.4E-10) and Toll-like receptor 7 (TLR7;  $\rho = 0.75$ , FDR = 2.4E-09). The top 2% ERC list of PDYN shows significant enrichment for terms related to immune system function (FDR = 6.0E-03), the complement and coagulation cascades (FDR = 6.0E-03), but no significant terms related to brain function other than “NCAM1 interactions” (FDR= 4.9E-02).

*IQ motif containing D (IQCD)*: IQCD is the 20<sup>th</sup> ranking ERC for ACE2 ( $\rho = 0.56$ , FDR = 8.9E-04). IQCD in mammals is not well studied. But it has been characterized as being involved in the “acrosome” (Zhang et al., 2019). The acrosome is an organelle that is part of the sperm and is involved in the “acrosome reaction”, which allows sperm to fuse with an egg upon fertilization (Abou-Haila & Tulsiani, 2000). It is required for spermatogenesis in mice (Harris et al., 2014) IQCD is therefore another protein with strong ERC to ACE2 implicated in male sex organs. There is also some evidence that suggests IQCD is associated with male fertility (Zhang et al., 2019). Additionally, ACE2 presence may be negatively associated with the acrosome reaction in sperm-precursor cells (Wang & Xu, 2020), but the direct mechanism for this is unclear. The top 2% ERC list for IQCD shows enrichment for tumor-necrosis factor-related terms (FDR = 9.3E-04) and “SW-620 cell” (4.9E-02) which is a human colon carcinoma cell line.

## J. Supplementary Text References

Aase K, Ernvist M, Ebarasi L, Jakobsson L, Majumdar A, Yi C, Birot O, Ming Y, Kvanta A, Edholm D, Aspenström P, Kissil J, Claesson-Welsh L, Shimono A, Holmgren L. 2007. Angiomotin regulates endothelial cell migration during embryonic angiogenesis. *Genes and Development* 21:2055–2068. DOI: 10.1101/gad.432007.

Abou-Haila A, Tulsiani DRP. 2000. Mammalian sperm acrosome: Formation, contents, and function. *Archives of Biochemistry and Biophysics* 379:173–182. DOI: 10.1006/abbi.2000.1880.

Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, Vanstapel A, Werlein C, Stark H, Tzankov A, Li WW, Li VW, Mentzer SJ, Jonigk D. 2020. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *New England Journal of Medicine* 383:120–128. DOI: 10.1056/nejmoa2015432.

Alsaid T, Aboulhosn JA, Cotts TB, Daniels CJ, Etheridge SP, Feltes TF, Gurvitz MZ, Lewin MB, Oster ME, Saidi A. 2020. Coronavirus Disease 2019 (COVID-19) Pandemic Implications in Pediatric and Adult Congenital Heart Disease. *Journal of the American Heart Association* 9:e017224. DOI: 10.1161/JAHA.120.017224.

Angiolini F, Cavallaro U. 2017. The pleiotropic role of L1CAM in tumor vasculature. *International Journal of Molecular Sciences* 18:254. DOI: 10.3390/ijms18020254.

Areschoug T, Gordon S. 2009. Scavenger receptors: Role in innate immunity and microbial pathogenesis. *Cellular Microbiology* 11:1160–1169. DOI: 10.1111/j.1462-5822.2009.01326.x.

Bader M. 2013. ACE2, angiotensin-(1-7), and Mas: The other side of the coin. *Pflugers Archiv European Journal of Physiology* 465:79–85. DOI: 10.1007/s00424-012-1120-0.

Balcar VJ, Zeman T, Janout V, Janoutová J, Lochman J, Šerý O. 2021. Single Nucleotide Polymorphism rs11136000 of CLU Gene (Clusterin, ApoJ) and the Risk of Late-Onset Alzheimer’s Disease in a Central European Population. *Neurochemical Research* 46:411–422. DOI: 10.1007/s11064-020-03176-y.

Bansal M. 2020. Cardiovascular disease and COVID-19. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews* 14:247–250. DOI: 10.1016/j.dsx.2020.03.013.

Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, Coetze R, Cukura A, Silva A Da, Denny P, Dogan T, Ebenezer TG, Fan J, Castro LG, Garmiri P, Georghiou G, Gonzales L, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Jokinen P, Joshi V, Jyothi D, Lock A, Lopez R, Luciani A, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Menchi M, Mishra A, Moulang K, Nightingale A, Oliveira CS, Pundir S, Qi G, Raj S, Rice D, Lopez MR, Saidi R, Sampson J, Sawford T, Speretta E, Turner E, Tyagi N, Vasudev P, Volynkin V, Warner K, Watkins X, Zaru R, Zellner H, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Bolleman J, Boutet E, Breuza L, Casals-Casas C, de Castro E, Echioukh KC, Coudert E, Cuche B, Doche M, Dornevil D, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hyka-Nouspikel N,

Jungo F, Keller G, Kerhornou A, Lara V, Le Mercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto TB, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Pozzato M, Pruess M, Rivoire C, Sigrist C, Sonesson K, Stutz A, Sundaram S, Tognolli M, Verbregue L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Zhang J. 2021. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research* 49:D480–D489. DOI: 10.1093/nar/gkaa1100.

Bernardes JP, Mishra N, Tran F, Bahmer T, Best L, Blase JI, Bordoni D, Franzenburg J, Geisen U, Josephs-Spaulding J, Köhler P, Künstner A, Rosati E, Aschenbrenner AC, Bacher P, Baran N, Boysen T, Brandt B, Bruse N, Dörr J, Dräger A, Elke G, Ellinghaus D, Fischer J, Forster M, Franke A, Franzenburg S, Frey N, Friedrichs A, Fuß J, Glück A, Hamm J, Hinrichsen F, Hoeppner MP, Imm S, Junker R, Kaiser S, Kan YH, Knoll R, Lange C, Laue G, Lier C, Lindner M, Marinos G, Markewitz R, Nattermann J, Noth R, Pickkers P, Rabe KF, Renz A, Röcken C, Rupp J, Schaffarzyk A, Scheffold A, Schulte-Schrepping J, Schunk D, Skowasch D, Ulas T, Wandinger KP, Wittig M, Zimmermann J, Busch H, Hoyer BF, Kaleta C, Heyckendorf J, Kox M, Rybníkář J, Schreiber S, Schultze JL, Rosenstiel P, Banovich NE, Desai T, Eickelberg O, Haniffa M, Horvath P, Kropski JA, Lafyatis R, Lundeberg J, Meyer K, Nawijn MC, Nikolic M, Ordovas Montanes J, Pe'er D, Tata PR, Rawlins E, Regev A, Reyfman P, Samakovlis C, Schultze J, Shalek A, Shepherd D, Spence J, Teichmann S, Theis F, Tsankov A, van den Berge M, von Papen M, Whitsett J, Zaragosi LE, Angelov A, Bals R, Bartholomäus A, Becker A, Bezdan D, Bonifacio E, Bork P, Clavel T, Colme-Tatché M, Diefenbach A, Dilthey A, Fischer N, Förstner K, Frick JS, Gagneur J, Goesmann A, Hain T, Hummel M, Janssen S, Kalinowski J, Kallies R, Kehr B, Keller A, Kim-Hellmuth S, Klein C, Kohlbacher O, Korbel JO, Kurth I, Landthaler M, Li Y, Ludwig K, Makarewicz O, Marz M, McHardy A, Mertes C, Nöthen M, Nürnberg P, Ohler U, Ossowski S, Overmann J, Peter S, Pfeffer K, Poetsch AR, Pühler A, Rajewsky N, Ralser M, Rieß O, Ripke S, Nunes da Rocha U, Saliba AE, Sander LE, Sawitzki B, Schiffer P, Schulte EC, Sczyrba A, Stegle O, Stoye J, Vehreschild J, Vogel J, von Kleist M, Walker A, Walter J, Wieczorek D, Ziebuhr J. 2020. Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. *Immunity* 53:1296–1314.e9. DOI: 10.1016/j.immuni.2020.11.017.

Bratt A, Birot O, Sinha I, Veitonmäki N, Aase K, Ernkvist M, Holmgren L. 2005. Angiomotin regulates endothelial cell-cell junctions and cell motility. *Journal of Biological Chemistry* 280:34859–34869. DOI: 10.1074/jbc.M503915200.

Buzza M, Wang YY, Dagher H, Babon JJ, Cotton RG, Powell H, Dowling J, Savige J. 2001. COL4A4 mutation in thin basement membrane disease previously described in Alport syndrome. *Kidney International* 60:480–483. DOI: 10.1046/j.1523-1755.2001.060002480.x.

Casino P, Gozalbo-Rovira R, Rodríguez-Díaz J, Banerjee S, Boutaud A, Rubio V, Hudson BG, Saus J, Cervera J, Marina A. 2018. Structures of collagen IV globular domains: Insight into associated pathologies, folding and network assembly. *IUCrJ* 5:765–779. DOI: 10.1107/S2052252518012459.

Caubit X, Lye CM, Martin E, Coré N, Long DA, Vola C, Jenkins D, Garratt AN, Skaer H, Woolf AS, Fasano L. 2008. Teashirt 3 is necessary for ureteral smooth muscle differentiation

downstream of SHH and BMP4. *Development* 135:3301–3310. DOI: 10.1242/dev.022442.

Caubit X, Thoby-Brisson M, Voituron N, Filippi P, Bévengut M, Faralli H, Zanella S, Fortin G, Hilaire G, Fasano L. 2010. Teashirt 3 regulates development of neurons involved in both respiratory rhythm and airflow control. *Journal of Neuroscience* 30:9465–9476. DOI: 10.1523/JNEUROSCI.1765-10.2010.

Chan YW, West S. 2015. GEN1 promotes Holliday junction resolution by a coordinated nick and counter-nick mechanism. *Nucleic Acids Research* 43:10882–10892. DOI: 10.1093/nar/gkv1207.

Chen X-L, Qiu L, Wang F, Liu S. 2014. Current understanding of tyrosine kinase BMX in inflammation and its inhibitors. *Burns & Trauma* 2:121. DOI: 10.4103/2321-3868.135483.

Chen YY, Zhang P, Zhou XM, Liu D, Zhong JC, Zhang CJ, Jin LJ, Yu HM. 2018. Relationship between genetic variants of ACE2 gene and circulating levels of ACE2 and its metabolites. *Journal of Clinical Pharmacy and Therapeutics* 43:189–195. DOI: 10.1111/jcpt.12625.

Chuang KH, Altuwaijri S, Li G, Lai JJ, Chu CY, Lai KP, Lin HY, Hsu JW, Keng P, Wu MC, Chang C. 2009. Neutropenia with impaired host defense against microbial infection in mice lacking androgen receptor. *Journal of Experimental Medicine* 206:1181–1199. DOI: 10.1084/jem.20082521.

Clarke TK, Ambrose-Lanci L, Ferraro TN, Berrettini WH, Kampman KM, Dackis CA, Pettinati HM, O'Brien CP, Oslin DW, Lohoff FW. 2012. Genetic association analyses of PDYN polymorphisms with heroin and cocaine addiction. *Genes, Brain and Behavior* 11:415–423. DOI: 10.1111/j.1601-183X.2012.00785.x.

Cogné B, Latypova X, Senaratne LDS, Martin L, Koboldt DC, Kellaris G, Fievet L, Le Meur G, Caldari D, Debray D, Nizon M, Frengen E, Bowne SJ, Buckley RM, Aberdein D, Alves PC, Barsh GS, Bellone RR, Bergström TF, Boyko AR, Brockman JA, Casal ML, Castelhano MG, Distl O, Dodman NH, Ellinwood NM, Fogle JE, Forman OP, Garrick DJ, Ginns EI, Häggström J, Harvey RJ, Hasegawa D, Haase B, Helps CR, Hernandez I, Hytönen MK, Kaukonen M, Kaelin CB, Kosho T, Leclerc E, Lear TL, Leeb T, Li RHL, Lohi H, Longeri M, Magnuson MA, Malik R, Mane SP, Munday JS, Murphy WJ, Pedersen NC, Rothschild MF, Rusbridge C, Shapiro B, Stern JA, Swanson WF, Terio KA, Todhunter RJ, Warren WC, Wilcox EA, Wildschutte JH, Yu Y, Cadena EL, Daiger SP, Bujakowska KM, Pierce EA, Gorin M, Katsanis N, Bézieau S, Petersen-Jones SM, Occelli LM, Lyons LA, Legeai-Mallet L, Sullivan LS, Davis EE, Isidor B. 2020. Mutations in the Kinesin-2 Motor KIF3B Cause an Autosomal-Dominant Ciliopathy. *American Journal of Human Genetics* 106:893–904. DOI: 10.1016/j.ajhg.2020.04.005.

Cramer TJ, Gale AJ. 2012. The anticoagulant function of coagulation factor V. *Thrombosis and Haemostasis* 107:15–21. DOI: 10.1160/TH11-06-0431.

Crocco P, Saiardi A, Wilson MS, Maletta R, Bruni AC, Passarino G, Rose G. 2016. Contribution of polymorphic variation of inositol hexakisphosphate kinase 3 (IP6K3) gene promoter to the susceptibility to late onset Alzheimer's disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1862:1766–1773. DOI: 10.1016/j.bbadi.2016.06.014.

Cunningham RL, Lumia AR, McGinnis MY. 2012. Androgen receptors, sex behavior, and

aggression. *Neuroendocrinology* 96:131–140. DOI: 10.1159/000337663.

Danziger RS. 2008. Aminopeptidase N in arterial hypertension. *Heart Failure Reviews* 13:293–298. DOI: 10.1007/s10741-007-9061-y.

Dolan ME, Hill DP, Mukherjee G, McAndrews MS, Chesler EJ, Blake JA. 2020. Investigation of COVID-19 comorbidities reveals genes and pathways coincident with the SARS-CoV-2 viral disease. *Scientific Reports* 10:1–11. DOI: 10.1038/s41598-020-77632-8.

Eisenreich A, Orphal M, Böhme K, Kreutz R. 2020. Tmem63c is a potential pro-survival factor in angiotensin II-treated human podocytes. *Life Sciences* 258. DOI: 10.1016/j.lfs.2020.118175.

Evans CE, Miners JS, Piva G, Willis CL, Heard DM, Kidd EJ, Good MA, Kehoe PG. 2020. ACE2 activation protects against cognitive decline and reduces amyloid pathology in the Tg2576 mouse model of Alzheimer's disease. *Acta Neuropathologica* 139:485–502. DOI: 10.1007/s00401-019-02098-6.

Fricke-Galindo I, Falfán-Valencia R. 2021. Genetics Insight for COVID-19 Susceptibility and Severity: A Review. *Frontiers in Immunology* 12. DOI: 10.3389/fimmu.2021.622176.

Gao M, Rendtlew Danielsen J, Wei LZ, Zhou DP, Xu Q, Li MM, Wang ZQ, Tong WM, Yang YG. 2012. A Novel Role of Human Holliday Junction Resolvase GEN1 in the Maintenance of Centrosome Integrity. *PLoS ONE* 7:e49687. DOI: 10.1371/journal.pone.0049687.

Gari K, Ortiz AML, Borel V, Flynn H, Skehel JM, Boulton SJ. 2012. MMS19 links cytoplasmic iron-sulfur cluster assembly to DNA metabolism. *Science* 337:243–245. DOI: 10.1126/science.1219664.

Garland T, Harvey PH, Ives AR. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 41:18–32. DOI: 10.1093/sysbio/41.1.18.

Gelmann EP. 2002. Molecular biology of the androgen receptor. *Journal of Clinical Oncology* 20:3001–3015. DOI: 10.1200/JCO.2002.10.018.

Gould WR, Silveira JR, Tracy PB. 2004. Unique in vivo modifications of coagulation factor V produce a physically and functionally distinct platelet-derived cofactor: Characterization of purified platelet-derived factor V/Va. *Journal of Biological Chemistry* 279:2383–2393. DOI: 10.1074/jbc.M308600200.

Grewal IS, Flavell RA. 1998. CD40 and CD154 in cell-mediated immunity. *Annual Review of Immunology* 16:111–135. DOI: 10.1146/annurev.immunol.16.1.111.

Handel MA, Schimenti JC. 2010. Genetics of mammalian meiosis: Regulation, dynamics and impact on fertility. *Nature Reviews Genetics* 11:124–136. DOI: 10.1038/nrg2723.

Harris TP, Schimenti KJ, Munroe RJ, Schimenti JC. 2014. IQ motif-containing G (Iqcg) is required for mouse spermiogenesis. *G3: Genes, Genomes, Genetics* 4:367–372. DOI: 10.1534/g3.113.009563.

Henriksson R, Bäckman CM, Harvey BK, Kadyrova H, Bazov I, Shippenberg TS, Bakalkin G.

2014. PDYN, a gene implicated in brain/mental disorders, is targeted by REST in the adult human brain. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1839:1226–1232. DOI: 10.1016/j.bbagr.2014.09.001.

Hodges AK, Piers TM, Collier D, Cousins O, Pocock JM. 2021. Pathways linking Alzheimer's disease risk genes expressed highly in microglia. *Neuroimmunology and Neuroinflammation* 2020. DOI: 10.20517/2347-8659.2020.60.

Hoxha M. 2020. What about COVID-19 and arachidonic acid pathway? *European Journal of Clinical Pharmacology* 76:1501–1504. DOI: 10.1007/s00228-020-02941-w.

Huang C, Ji X, Zhou W, Huang Z, Peng X, Fan L, Lin G, Zhu W. 2021. Coronavirus: A possible cause of reduced male fertility. *Andrology* 9:80–87. DOI: 10.1111/andr.12907.

Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution* 33:1635–1638. DOI: 10.1093/molbev/msw046.

Hurtado-Guerrero I, Hernández B, Pinto-Medel MJ, Calonge E, Rodriguez-Bada JL, Urbaneja P, Alonso A, Mena-Vázquez N, Aliaga P, Issazadeh-Navikas S, Pavia J, Leyva L, Alcamí J, Alcamí A, Fernández Ó, Oliver-Martos B. 2020. Antiviral, Immunomodulatory and Antiproliferative Activities of Recombinant Soluble IFNAR2 without IFN-β Mediation. *Journal of Clinical Medicine* 9:959. DOI: 10.3390/jcm9040959.

Itakura E, Chiba M, Murata T, Matsuura A. 2020. Heparan sulfate is a clearance receptor for aberrant extracellular proteins. *Journal of Cell Biology* 219. DOI: 10.1083/JCB.201911126.

Ivanciu L, Crosby J, MacLeod AR, Revenko A, Camire RM, Davidson RJ, Monia BP. 2017. Differential role of plasma and platelet-derived factor V in vivo. *Blood* 130:364–364. DOI: 10.1182/BLOOD.V130.SUPPL\_1.364.364.

Jamin SP, Petit FG, Demini L, Primig M. 2021. Tex55 encodes a conserved putative A-kinase anchoring protein dispensable for male fertility in the mouse. *Biology of Reproduction* 104:731–733. DOI: 10.1093/biolre/ioab007.

Jenne DE, Tschopp J. 1989. Molecular structure and functional characterization of a human complement cytosis inhibitor found in blood and seminal plasma: Identity to sulfated glycoprotein 2, a constituent of rat testis fluid. *Proceedings of the National Academy of Sciences of the United States of America* 86:7123–7127. DOI: 10.1073/pnas.86.18.7123.

Jeziorska J, Stevanin G, Watanabe H, Fokkens MR, Zagnoli F, Kok J, Goas JY, Bertrand P, Robin C, Brice A, Bakalkin G, Durr A, Verbeek DS. 2013. Identification and characterization of novel PDYN mutations in dominant cerebellar ataxia cases. *Journal of Neurology* 260:1807–1812. DOI: 10.1007/s00415-013-6882-6.

Jin HY, Chen LJ, Zhang ZZ, Xu Y Le, Song B, Xu R, Oudit GY, Gao PJ, Zhu DL, Zhong JC. 2015. Deletion of angiotensin-converting enzyme 2 exacerbates renal inflammation and injury in apolipoprotein E-deficient mice through modulation of the nephrin and TNF-alpha-TNFRSF1A signaling. *Journal of Translational Medicine* 13:1–16. DOI: 10.1186/s12967-015-0616-8.

Joseph NF, Grinman E, Swarnkar S, Puthanveettil S V. 2020. Molecular Motor KIF3B Acts as a

Key Regulator of Dendritic Architecture in Cortical Neurons. *Frontiers in Cellular Neuroscience* 14:1–11. DOI: 10.3389/fncel.2020.521199.

Jung SM, Namgoong S, Seo JM, Kim DY, Oh JT, Kim HY, Kim JH. 2019. Potential association between TSGA13 variants and risk of total colonic aganglionosis in Hirschsprung disease. *Gene* 710:240–245. DOI: 10.1016/j.gene.2019.06.007.

Kai H, Kai M. 2020. Interactions of coronaviruses with ACE2, angiotensin II, and RAS inhibitors—lessons from available evidence and insights into COVID-19. *Hypertension Research* 43:648–654. DOI: 10.1038/s41440-020-0455-8.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28:27–30. DOI: 10.1093/nar/28.1.27.

Kawai T, Akira S. 2010. The role of pattern-recognition receptors in innate immunity: Update on toll-like receptors. *Nature Immunology* 11:373–384. DOI: 10.1038/ni.1863.

Kehoe PG. 2018. The coming of age of the angiotensin hypothesis in Alzheimer's disease: Progress toward disease prevention and treatment? *Journal of Alzheimer's Disease* 62:1443–1466. DOI: 10.3233/JAD-171119.

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464. DOI: 10.1093/bioinformatics/btq166.

Kim S. 2015. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods* 22:665–674. DOI: 10.5351/csam.2015.22.6.665.

Van Kooten G, Banchereau J. 2000. CD40-CD40 ligand. *Journal of Leukocyte Biology* 67:2–17. DOI: 10.1002/jlb.67.1.2.

Kriventseva E V., Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* 47:D807–D811. DOI: 10.1093/nar/gky1053.

Kuba K, Imai Y, Ohto-Nakanishi T, Penninger JM. 2010. Trilogy of ACE2: A peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. *Pharmacology and Therapeutics* 128:119–128. DOI: 10.1016/j.pharmthera.2010.06.003.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular biology and evolution* 34:1812–1819. DOI: 10.1093/molbev/msx116.

Lambert DW, Yarski M, Warner FJ, Thornhill P, Parkin ET, Smith AI, Hooper NM, Turner AJ. 2005. Tumor necrosis factor- $\alpha$  convertase (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2 (ACE2). *Journal of Biological Chemistry* 280:30113–30119. DOI: 10.1074/jbc.M505111200.

Lei Y, Takahama Y. 2012. XCL1 and XCR1 in the immune system. *Microbes and Infection*

14:262–267. DOI: 10.1016/j.micinf.2011.10.003.

Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*. DOI: 10.1093/nar/gkab301.

Lim B, Sando SB, Grøntvedt GR, Bråthen G, Diamandis EP. 2020. Cerebrospinal fluid neuronal pentraxin receptor as a biomarker of long-term progression of Alzheimer's disease: a 24-month follow-up study. *Neurobiology of Aging* 93:97.e1-97.e7. DOI: 10.1016/j.neurobiolaging.2020.03.013.

Liu X, Chen Y, Tang W, Zhang L, Chen W, Yan Z, Yuan P, Yang M, Kong S, Yan L, Qiao J. 2020. Single-cell transcriptome analysis of the novel coronavirus (SARS-CoV-2) associated gene ACE2 expression in normal and non-obstructive azoospermia (NOA) human male testes. *Science China Life Sciences* 63:1006–1015. DOI: 10.1007/s11427-020-1705-0.

Liu D, Yang J, Feng B, Lu W, Zhao C, Li L. 2021. Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19. *Journal of Infection* 82:126–132. DOI: 10.1016/j.jinf.2020.11.031.

Longo I, Porcedda P, Mari F, Giachino D, Meloni I, Deplano C, Brusco A, Bosio M, Massella L, Lavoratti G, Roccatello D, Frascá G, Mazzucco G, Onetti Muda A, Conti M, Fasciolo F, Arrondel C, Heidet L, Renieri A, De Marchi M. 2002. COL4A3/COL4A4 mutations: From familial hematuria to autosomal-dominant or recessive Alport syndrome. *Kidney International* 61:1947–1956. DOI: 10.1046/j.1523-1755.2002.00379.x.

Louwersheimer E, Cohn-Hokke PE, Pijnenburg YAL, Weiss MM, Sistermans EA, Rozemuller AJ, Hulsman M, Van Swieten JC, Van Duijn CM, Barkhof F, Koenei T, Scheltens P, Van Der Flier WM, Holstege H. 2017. Rare genetic variant in SORL1 may increase penetrance of Alzheimer's disease in a family with several generations of APOE-ε4 homozygosity. *Journal of Alzheimer's Disease* 56:63–74. DOI: 10.3233/JAD-160091.

Lu YC, Xavier-Ferrucio J, Wang L, Zhang PX, Grimes HL, Venkatasubramanian M, Chetal K, Aronow B, Salomonis N, Krause DS. 2018. The Molecular Signature of Megakaryocyte-Erythroid Progenitors Reveals a Role for the Cell Cycle in Fate Specification. *Cell Reports* 25:2083-2093.e4. DOI: 10.1016/j.celrep.2018.10.084.

Maglott D, Ostell J, Pruitt KD, Tatusova T. 2005. Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Research* 33:D54–D58. DOI: 10.1093/nar/gki031.

Matsumoto T, Shiina H, Kawano H, Sato T, Kato S. 2008. Androgen receptor functions in male and female physiology. *Journal of Steroid Biochemistry and Molecular Biology* 109:236–241. DOI: 10.1016/j.jsbmb.2008.03.023.

McKechnie JL, Blish CA. 2020. The Innate Immune System: Fighting on the Front Lines or Fanning the Flames of COVID-19? *Cell Host and Microbe* 27:863–869. DOI: 10.1016/j.chom.2020.05.009.

Miyata H, Castaneda JM, Fujihara Y, Yu Z, Archambeault DR, Isotani A, Kiyozumi D, Kriseman ML, Mashiko D, Matsumura T, Matzuk RM, Mori M, Noda T, Oji A, Okabe M, Prunskaitė-Hyyryläinen R, Ramirez-Solis R, Satouh Y, Zhang Q, Ikawa M, Matzuk MM. 2016. Genome engineering uncovers 54 evolutionarily conserved and testis-enriched genes

that are not required for male fertility in mice. *Proceedings of the National Academy of Sciences of the United States of America* 113:7704–7710. DOI: 10.1073/pnas.1608458113.

Moos M, Tacke R, Scherer H, Teplow D, Früh K, Schachner M. 1988. Neural adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin. *Nature* 334:701–703. DOI: 10.1038/334701a0.

Mosesson MW. 2005. Fibrinogen and fibrin structure and functions. In: *Journal of Thrombosis and Haemostasis*. John Wiley & Sons, Ltd, 1894–1904. DOI: 10.1111/j.1538-7836.2005.01365.x.

Nauen DW, Hooper JE, Stewart CM, Solomon IH. 2021. Assessing Brain Capillaries in Coronavirus Disease 2019. *JAMA Neurology*. DOI: 10.1001/jamaneurol.2021.0225.

Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, Powell CC, Nassar LR, Maulding ND, Lee CM, Lee BT, Hinrichs AS, Fyfe AC, Fernandes JD, Diekhans M, Clawson H, Casper J, Benet-Pagès A, Barber GP, Haussler D, Kuhn RM, Haeussler M, Kent WJ. 2021. The UCSC genome browser database: 2021 update. *Nucleic Acids Research* 49:D1046–D1057. DOI: 10.1093/nar/gkaa1070.

Novikova G, Kapoor M, Tew J, Abud EM, Efthymiou AG, Chen SX, Cheng H, Fullard JF, Bendl J, Liu Y, Roussos P, Björkegren JL, Liu Y, Poon WW, Hao K, Marcora E, Goate AM. 2021. Integration of Alzheimer’s disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nature Communications* 12:1–14. DOI: 10.1038/s41467-021-21823-y.

Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH, Russell CD, Furniss J, Richmond A, Gountouna E, Wrobel N, Harrison D, Wang B, Wu Y, Meynert A, Griffiths F, Oosthuyzen W, Kousathanas A, Moutsianas L, Yang Z, Zhai R, Zheng C, Grimes G, Beale R, Millar J, Shih B, Keating S, Zechner M, Haley C, Porteous DJ, Hayward C, Yang J, Knight J, Summers C, Shankar-Hari M, Klenerman P, Turtle L, Ho A, Moore SC, Hinds C, Horby P, Nichol A, Maslove D, Ling L, McAuley D, Montgomery H, Walsh T, Pereira AC, Renieri A, Shen X, Ponting CP, Fawkes A, Tenesa A, Caulfield M, Scott R, Rowan K, Murphy L, Openshaw PJM, Semple MG, Law A, Vitart V, Wilson JF, Baillie JK. 2021. Genetic mechanisms of critical illness in COVID-19. *Nature* 591:92–98. DOI: 10.1038/s41586-020-03065-y.

Palmer N, Kaldis P. 2020. Less-well known functions of cyclin/CDK complexes. *Seminars in Cell and Developmental Biology* 107:54–62. DOI: 10.1016/j.semcdcb.2020.04.003.

Paradis E, Schliep K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528. DOI: 10.1093/bioinformatics/bty633.

Peng X, Xu E, Liang W, Pei X, Chen D, Zheng D, Zhang Y, Zheng C, Wang P, She S, Zhang Y, Ma J, Mo X, Zhang Y, Ma D, Wang Y. 2016. Identification of FAM3D as a new endogenous chemotaxis agonist for the formyl peptide receptors. *Journal of Cell Science* 129:1831–1842. DOI: 10.1242/jcs.183053.

Prokunina-Olsson L, Alphonse N, Dickenson RE, Durbin JE, Glenn JS, Hartmann R, Kotenko S V., Lazear HM, O’Brien TR, Odendall C, Onabajo OO, Piontkevska H, Santer DM, Reich

NC, Wack A, Zanoni I. 2020. COVID-19 and emerging viral infections: The case for interferon lambda. *Journal of Experimental Medicine* 217. DOI: 10.1084/jem.20200653.

Rangel R, Sun Y, Guzman-Rojas L, Ozawa MG, Sun J, Giordano RJ, Van Pelt CS, Tinkey PT, Behringer RR, Sidman RL, Arap W, Pasqualini R. 2007. Impaired angiogenesis in aminopeptidase N-null mice. *Proceedings of the National Academy of Sciences of the United States of America* 104:4588–4593. DOI: 10.1073/pnas.0611653104.

Reed AAC, Loh NY, Terryn S, Lippiat JD, Partridge C, Galvanovskis J, Williams SE, Jouret F, Wu FTF, Courtoy PJ, Andrew Nesbit M, Rorsman P, Devuyst O, Ashcroft FM, Thakker R V. 2010. CLC-5 and KIF3B interact to facilitate CLC-5 plasma membrane expression, endocytosis, and microtubular transport: Relevance to pathophysiology of Dent's disease. *American Journal of Physiology - Renal Physiology* 298:F365–F380. DOI: 10.1152/ajprenal.00038.2009.

Ripon MAR, Bhowmik DR, Amin MT, Hossain MS. 2021. Role of arachidonic cascade in COVID-19 infection: A review. *Prostaglandins and Other Lipid Mediators* 154:106539. DOI: 10.1016/j.prostaglandins.2021.106539.

Rizvanov AA, Kiyasov AP, Gazizov IM, Yilmaz TS, Kaligin MS, Andreeva DI, Shafigullina AK, Guseva DS, Kiselev SL, Matin K, Palotás A, Islamov RR. 2008. Human umbilical cord blood cells transfected with VEGF and L1CAM do not differentiate into neurons but transform into vascular endothelial cells and secrete neuro-trophic factors to support neurogenesis-a novel approach in stem cell therapy. *Neurochemistry International* 53:389–394. DOI: 10.1016/j.neuint.2008.09.011.

Samatov TR, Wicklein D, Tonevitsky AG. 2016. L1CAM: Cell adhesion and more. *Progress in Histochemistry and Cytochemistry* 51:25–32. DOI: 10.1016/j.proghi.2016.05.001.

Samuel RM, Majd H, Richter MN, Ghazizadeh Z, Zekavat SM, Navickas A, Ramirez JT, Asgharian H, Simoneau CR, Bonser LR, Koh KD, Garcia-Knight M, Tassetto M, Sunshine S, Farahvashi S, Kalantari A, Liu W, Andino R, Zhao H, Natarajan P, Erle DJ, Ott M, Goodarzi H, Fattah F. 2020. Androgen Signaling Regulates SARS-CoV-2 Receptor Levels and Is Associated with Severe COVID-19 Symptoms in Men. *Cell Stem Cell* 27:876–889.e12. DOI: 10.1016/j.stem.2020.11.009.

Sánchez-Martín P, Komatsu M. 2020. Heparan sulfate and clusterin: Cleaning squad for extracellular protein degradation. *Journal of Cell Biology* 219. DOI: 10.1083/JCB.202001159.

Schäfer MKE, Altevogt P. 2010. L1CAM malfunction in the nervous system and human carcinomas. *Cellular and Molecular Life Sciences* 67:2425–2437. DOI: 10.1007/s00018-010-0339-1.

Schulz A, Müller NV, Van De Lest NA, Eisenreich A, Schmidbauer M, Barysenka A, Purfürst B, Sporbert A, Lorenzen T, Meyer AM, Herlan L, Witten A, Rühle F, Zhou W, De Heer E, Scharpfenecker M, Panáková D, Stoll M, Kreutz R. 2019. Analysis of the genomic architecture of a complex trait locus in hypertensive rat models links Tmem63C to kidney damage. *eLife* 8. DOI: 10.7554/eLife.42068.

Severe Covid-19 GWAS Group. 2020. Genomewide Association Study of Severe Covid-19 with

Respiratory Failure. *New England Journal of Medicine* 383:1522–1534. DOI: 10.1056/nejmoa2020283.

Seymen CM. 2021. The other side of COVID-19 pandemic: Effects on male fertility. *Journal of Medical Virology* 93:1396–1402. DOI: 10.1002/jmv.26667.

Shepardson KM, Larson K, Johns LL, Stanek K, Cho H, Wellham J, Henderson H, Rynda-Apple A. 2018. IFNAR2 is required for anti-influenza immunity and alters susceptibility to post-influenza bacterial superinfections. *Frontiers in Immunology* 9:2589. DOI: 10.3389/fimmu.2018.02589.

Shepherd CE, Affleck AJ, Bahar AY, Carew-Jones F, Halliday GM. 2020. Intracellular and secreted forms of clusterin are elevated early in Alzheimer's disease and associate with both A $\beta$  and tau pathology. *Neurobiology of Aging* 89:129–131. DOI: 10.1016/j.neurobiolaging.2019.10.025.

Singh MK, Mobeen A, Chandra A, Joshi S, Ramachandran S. 2021. A meta-analysis of comorbidities in COVID-19: Which diseases increase the susceptibility of SARS-CoV-2 infection? *Computers in Biology and Medicine* 130:104219. DOI: 10.1016/j.combiomed.2021.104219.

Srinivasan K, Friedman BA, Etxeberria A, Huntley MA, van der Brug MP, Foreman O, Paw JS, Modrusan Z, Beach TG, Serrano GE, Hansen D V. 2020. Alzheimer's Patient Microglia Exhibit Enhanced Aging and Unique Transcriptional Activation. *Cell Reports* 31:107843. DOI: 10.1016/j.celrep.2020.107843.

Stefely JA, Christensen BB, Gogakos T, Cone Sullivan JK, Montgomery GG, Barranco JP, Van Cott EM. 2020. Marked factor V activity elevation in severe COVID-19 is associated with venous thromboembolism. *American Journal of Hematology* 95:1522–1530. DOI: 10.1002/ajh.25979.

Stehling O, Vashisht AA, Mascarenhas J, Jonsson ZO, Sharma T, Netz DJA, Pierik AJ, Wohlschlegel JA, Lill R. 2012. MMS19 assembles iron-sulfur proteins required for DNA metabolism and genomic integrity. *Science* 337:195–199. DOI: 10.1126/science.1219723.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Iny Stein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. 2016. The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics* 2016:1.30.1–1.30.33. DOI: 10.1002/cpbi.5.

Sutton BS, Crosslin DR, Shah SH, Nelson SC, Bassil A, Hale AB, Haynes C, Goldschmidt-Clermont PJ, Vance JM, Seo D, Kraus WE, Gregory SG, Hauser ER. 2008. Comprehensive genetic analysis of the platelet activating factor acetylhydrolase (PLA2G7) gene and cardiovascular disease in case-control and family datasets. *Human Molecular Genetics* 17:1318–1328. DOI: 10.1093/hmg/ddn020.

Tabach Y, Golan T, Hernández-Hernández A, Messer AR, Fukuda T, Kouznetsova A, Liu JG, Lilenthal I, Levy C, Ruvkun G. 2013. Human disease locus discovery and mapping to molecular pathways through phylogenetic profiling. *Molecular Systems Biology* 9:692. DOI: 10.1038/msb.2013.50.

Takata N, Ishii K aki, Takayama H, Nagashimada M, Kamoshita K, Tanaka T, Kikuchi A, Takeshita Y, Matsumoto Y, Ota T, Yamamoto Y, Yamagoe S, Seki A, Sakai Y, Kaneko S, Takamura T. 2021. LECT2 as a hepatokine links liver steatosis to inflammation via activating tissue macrophages in NASH. *Scientific Reports* 11. DOI: 10.1038/s41598-020-80689-0.

Takeda K, Kaisho T, Akira S. 2003. Toll-like receptors. *Annual Review of Immunology* 21:335–376. DOI: 10.1146/annurev.immunol.21.120601.141126.

Thomas C, Moraga I, Levin D, Krutzik PO, Podoplelova Y, Trejo A, Lee C, Yarden G, Vleck SE, Glenn JS, Nolan GP, Piehler J, Schreiber G, Garcia KC. 2011. Structural linkage between ligand discrimination and receptor activation by Type i interferons. *Cell* 146:621–632. DOI: 10.1016/j.cell.2011.06.048.

Thul PJ, Akesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, Fall J, Gatto L, Gnann C, Hober S, Hjelmare M, Johansson F, Lee S, Lindskog C, Mulder J, Mulvey CM, Nilsson P, Oksvold P, Rockberg J, Schutten R, Schwenk JM, Sivertsson A, Sjöstedt E, Skogs M, Stadler C, Sullivan DP, Tegel H, Winsnes C, Zhang C, Zwahlen M, Mardinoglu A, Pontén F, Von Feilitzen K, Lilley KS, Uhlén M, Lundberg E. 2017. A subcellular map of the human proteome. *Science* 356. DOI: 10.1126/science.aal3321.

Tschopp J, Jenne DE, Hertig S, Preissner KT, Morgenstern H, Sapino - AP, French L. 1993. Human megakaryocytes express clusterin and package it without apolipoprotein A-1 into  $\alpha$ -granules. *Blood* 82:118–125. DOI: 10.1182/blood.v82.1.118.bloodjournal821118.

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson IM, Edlund K, Lundberg E, Navani S, Szigyarto CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, Von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, Von Heijne G, Nielsen J, Pontén F. 2015. Tissue-based map of the human proteome. *Science* 347:1260419–1260419. DOI: 10.1126/science.1260419.

Vainio P, Gupta S, Ketola K, Mirtti T, Mpindi JP, Kohonen P, Fey V, Perälä M, Smit F, Verhaegh G, Schalken J, Alanen KA, Kallioniemi O, Iljin K. 2011. Arachidonic acid pathway members PLA2G7, HPGD, EPHX2, and CYP4F8 identified as putative novel therapeutic targets in prostate cancer. *American Journal of Pathology* 178:525–536. DOI: 10.1016/j.ajpath.2010.10.002.

Verma S, Saksena S, Sadri-Ardekani H. 2020. ACE2 receptor expression in testes: Implications in coronavirus disease 2019 pathogenesis. *Biology of Reproduction* 103:449–451. DOI: 10.1093/biolre/ioaa080.

Vietri M, Stenmark H. 2018. Orchestrating Nuclear Envelope Sealing during Mitosis. *Developmental Cell* 47:541–542. DOI: 10.1016/j.devcel.2018.11.020.

Wang Q, Hao Y, Mo X, Wang L, Lu X, Huang J, Cao J, Li H, Gu D. 2010. PLA2G7 gene polymorphisms and coronary heart disease risk: A meta-analysis. *Thrombosis Research* 126:498–503. DOI: 10.1016/j.thromres.2010.09.009.

Wang Z, Xu X. 2020. scRNA-seq Profiling of Human Testes Reveals the Presence of the ACE2 Receptor, A Target for SARS-CoV-2 Infection in Spermatogonia, Leydig and Sertoli Cells. *Cells* 9:920. DOI: 10.3390/cells9040920.

de Wit NJW, IJssennagger N, Oosterink E, Keshtkar S, Hooiveld GJEJ, Mensink RP, Hammer S, Smit JWA, Müller M, der Meer R van. 2012. Oit1/Fam3D, a gut-secreted protein displaying nutritional status-dependent regulation. *Journal of Nutritional Biochemistry* 23:1425–1433. DOI: 10.1016/j.jnutbio.2011.09.003.

Witte DP, Aronow BJ, Stauderman ML, Stuart WD, Clay MA, Gruppo RA, Jenkins SH, Harmony JAK. 1993. Platelet activation releases megakaryocyte-synthesized apolipoprotein J, a highly abundant protein in atherosomatous lesions. *American Journal of Pathology* 143:763–773.

Yamada T, Nakao K, Itoh H, Morii N, Shiono S, Sakamoto M, Sugawara A, Saito Y, Mukoyama M, Arai H, Eigyo M, Matsushita A, Imura H. 1988. Inhibitory action of leumorphin on vasopressin secretion in conscious rats. *Endocrinology* 122:985–990. DOI: 10.1210/endo-122-3-985.

Yamagoe S, Yamakawa Y, Matsuo Y, Minowada J, Mizuno S, Suzuki K. 1996. Purification and primary amino acid sequence of a novel neutrophil chemotactic factor LECT2. *Immunology Letters* 52:9–13. DOI: 10.1016/0165-2478(96)02572-2.

Yamamoto K, Ishiai M, Matsushita N, Arakawa H, Lamerdin JE, Buerstedde J-M, Tanimoto M, Harada M, Thompson LH, Takata M. 2003. Fanconi Anemia FANCG Protein in Mitigating Radiation- and Enzyme-Induced DNA Double-Strand Breaks by Homologous Recombination in Vertebrate Cells. *Molecular and Cellular Biology* 23:5421–5430. DOI: 10.1128/mcb.23.15.5421-5430.2003.

Yeager CL, Ashmun RA, Williams RK, Cardellicchio CB, Shapiro LH, Look AT, Holmes K V. 1992. Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* 357:420–422. DOI: 10.1038/357420a0.

Zhang P, Jiang W, Luo N, Zhu W, Fan L. 2019. IQ motif containing D (IQCD), a new acrosomal protein involved in the acrosome reaction and fertilisation. *Reproduction, Fertility and Development*. DOI: 10.1071/RD18416.

Zhang Q, Liu Z, Moncada-Velez M, Chen J, Ogishi M, Bigio B, Yang R, Arias AA, Zhou Q, Han JE, Ugurbil AC, Zhang P, Rapaport F, Li J, Spaan AN, Boisson B, Boisson-Dupuis S, Bustamante J, Puel A, Ciancanelli MJ, Zhang SY, Béziat V, Jouanguy E, Abel L, Cobat A, Casanova JL, Bastard P, Korol C, Rosain J, Philippot Q, Chbihi M, Lorenzo L, Bizien L, Neehus AL, Kerner G, Seeleuthner Y, Manry J, Le Voyer T, Boisson B, Boisson-Dupuis S, Bustamante J, Puel A, Zhang SY, Béziat V, Jouanguy E, Abel L, Cobat A, Casanova JL, Bastard P, Rosain J, Philippot Q, Chbihi M, Lorenzo L, Bizien L, Neehus AL, Kerner G, Seeleuthner Y, Manry J, Le Voyer T, Boisson B, Boisson-Dupuis S, Bustamante J, Puel A, Zhang SY, Béziat V, Jouanguy E, Abel L, Cobat A, Casanova JL, Le Pen J, Schneider WM, Razook BS, Hoffmann HH, Michailidis E, Rice CM, Sabli IKD, Hodeib S, Sancho-Shimizu V, Bilguvar K, Ye J, Maniatis T, Bolze A, Arias AA, Arias AA, Zhang Y, Notarangelo LD, Su HC, Zhang Y, Notarangelo LD, Su HC, Onodi F, Korniotis S, Karpf L, Soumelis V, Bonnet-Madin L, Amara A, Dorgham K, Gorochov G, Smith N, Duffy D,

Moens L, Meyts I, Meade P, García-Sastre A, Krammer F, Corneau A, Masson C, Schmitt Y, Schlüter A, Pujol A, Khan T, Marr N, Fellay J, Fellay J, Roussel L, Vinh DC, Shahrooei M, Shahrooei M, Alosaimi MF, Alsohime F, Hasanato R, Mansouri D, Mansouri D, Mansouri D, Al-Saud H, Almourfi F, Al-Mulla F, Al-Muhsen SZ, Al Turki S, Al Turki S, van de Beek D, Biondi A, Bettini LR, D'Angio M, Bonfanti P, Imberti L, Sottini A, Paghera S, Quiros-Roldan E, Rossi C, Oler AJ, Tompkins MF, Alba C, Dalgard CL, Vandernoot I, Smits G, Goffard JC, Migeotte I, Haerynck F, Soler-Palacin P, Martin-Nalda A, Colobran R, Morange PE, Keles S, Çölkesen F, Ozcelik T, Yasar KK, Senoglu S, Karabela ŞN, Rodriguez-Gallego C, Rodríguez-Gallego C, Novelli G, Hraiech S, Tandjaoui-Lambotte Y, Tandjaoui-Lambotte Y, Duval X, Laouénan C, Duval X, Laouénan C, Laouénan C, Snow AL, Dalgard CL, Milner JD, Mogensen TH, Mogensen TH, Marr N, Spaan AN, Bustamante J, Ciancanelli MJ, Meyts I, Maniatis T, Soumelis V, Nussenzweig M, Nussenzweig M, Casanova JL, García-Sastre A, García-Sastre A, García-Sastre A, Lifton RP, Lifton RP, Lifton RP, Casanova JL, Foti G, Bellani G, Citerio G, Contro E, Pesci A, Valsecchi MG, Cazzaniga M, Abad J, Blanco I, Rodrigo C, Aguilera-Albesa S, Akcan OM, Darazam IA, Aldave JC, Ramos MA, Nadji SA, Alkan G, Allardet-Servent J, Allende LM, Alsina L, Alyanakian MA, Amador-Borrero B, Mouly S, Sene D, Amoura Z, Mathian A, Antolí A, Blanch GR, Riera JS, Moreno XS, Arslan S, Assant S, Auguet T, Azot A, Bajolle F, Bustamante J, Lévy R, Oualha M, Baldolli A, Ballester M, Feldman HB, Barrou B, Beurton A, Bilbao A, Blanchard-Rohner G, Blandinières A, Rivet N, Blazquez-Gamero D, Bloomfield M, Bolivar-Prados M, Clavé P, Borie R, Bosteels C, Lambrecht BN, van Braeckel E, Bousfiha AA, Bouvattier C, Vincent A, Boyarchuk O, Bueno MRP, Castro M V., Matos LRB, Zatz M, Agra JJC, Calimli S, Capra R, Carrabba M, Fabio G, Casasnovas C, Vélez-Santamaría V, Caseris M, Falck A, Poncelet G, Castelle M, Castelli F, de Vera MC, Catherinot E, Chalumeau M, Toubiana J, Charbit B, Li Z, Pellegrini S, Cheng MP, Clotet B, Codina A, Colkesen F, Çölkesen F, Colobran R, Comarmond C, Dalmau D, Dalmau D, Darley DR, Dauby N, Dauger S, Le Bourgeois F, Levy M, de Pontual L, Dehban A, Delplancq G, Demoule A, Diehl JL, Dobbelaere S, Durand S, Mircher C, Rebillat AS, Vilaire ME, Eldars W, Elgamal M, Elnagdy MH, Emiroglu M, Erdeniz EH, Aytekin SE, Euvrard R, Evcen R, Faivre L, Fartoukh M, Philippot Q, Faure M, Arquero MF, Flores C, Flores C, Flores C, Francois B, Fumadó V, Fumadó V, Fumadó V, Fusco F, Ursini MV, Solis BG, de Diego RP, van Den Rym AM, Gaussem P, Gil-Herrera J, Gilardin L, Alarcon MG, Girona-Alarcón M, Goffard JC, Gok F, Yosunkaya A, González-Montelongo R, Íñigo-Campos A, Lorenzo-Salazar JM, Muñoz-Barrera A, Guerder A, Gul Y, Guner SN, Gut M, Hadjadj J, Haerynck F, Halwani R, Hammarström L, Hatipoglu N, Hernandez-Brito E, Heijmans C, Holanda-Peña MS, Horcajada JP, Hoste L, Hoste E, Hraiech S, Humbert L, Mordacq C, Thumerelle C, Vuotto F, Iglesias AD, Jamme M, Arranz MJ, Jordan I, Jorens P, Kanat F, Kapakli H, Kara I, Karbuz A, Yasar KK, Senoglu S, Keles S, Demirkol YK, Klocperk A, Król ZJ, Kuentz P, Kwan YWM, Lagier JC, Lau YL, Leung D, Leo YS, Young BE, Lopez RL, Levin M, Linglart A, Loeys B, Louapre C, Lubetzki C, Luyt CE, Lye DC, Mansouri D, Marjani M, Pereira JM, Martin A, Soler-Palacín P, Pueyo DM, Martinez-Picado J, Marzana I, Matthews G V., Mayaux J, Parizot C, Quentrec P, Mège JL, Raoult D, Melki I, Meritet JF, Metin O, Meyts I, Mezidi M, Migeotte I, Taccone F, Millereux M, Mirault T, Mirsaeidi M, Melián AM, Martinez AM, Morange P, Morelle G, Naesens L, Nafati C, Neves JF, Ng LFP, Medina YN, Cuadros EN, Gonzalo Ocejo-Vinyals J, Orbak Z, Özçelik T, Pan-

Hammarström Q, Pascreau T, Paz-Artal E, Philippe A, Planas-Serra L, Schluter A, Ploin D, Viel S, Poissy J, Pouletty M, Reisli I, Ricart P, Richard JC, Rivière JG, Rodriguez-Gallego C, Rodriguez-Gallego C, Rodríguez-Palmero A, Romero CS, Rothenbuhler A, Rozenberg F, del Prado MYR, Sanchez O, Sánchez-Ramón S, Schmidt M, Schweitzer CE, Scolari F, Sediva A, Seijo LM, Seppänen MRJ, Ilovich AS, Slabbynck H, Smadja DM, Sobh A, Solé-Violán J, Soler C, Stepanovskiy Y, Stoclin A, Tandjaoui-Lambiotte Y, Taupin JL, Tavernier SJ, Terrier B, Tomasoni G, Alvarez JT, Trouillet-Assant S, Troya J, Tucci A, Uzunhan Y, Vabres P, Valencia-Ramos J, van de Velde S, van Praet J, Vandernoot I, Vatansev H, Vilain C, Voiriot G, Yucel F, Zannad F, Belot A, Bole-Feysot C, Lyonnet S, Masson C, Nitschke P, Pouliet A, Schmitt Y, Tores F, Zarhrate M, Shahrooei M, Abel L, Andrejak C, Angoulvant F, Bachelet D, Bhavsar K, Bouadma L, Chair A, Couffignal C, Silveira C Da, Debray MP, Duval X, Eloy P, Esposito-Farese M, Ettalhaoui N, Gault N, Ghosn J, Gorenne I, Hoffmann I, Kafif O, Kali S, Khalil A, Laouénan C, Laribi S, Le M, Le Hingrat Q, Lescure FX, Lucet JC, Mentré F, Mullaert J, Peiffer-Smadja N, Peytavin G, Roy C, Schneider M, Mohammed NS, Tagherset L, Tardivon C, Tellier MC, Timsit JF, Trioux T, Tubiana S, Basmaci R, Behillil S, Beluze M, Benkerrou D, Dorival C, Meziane A, Téoulé F, Bompard F, Bouscambert M, Gaymard A, Lina B, Rosa-Calatrava M, Terrier O, Caralp M, Cervantes-Gonzalez M, D'Ortenzio E, Puéchal O, Semaille C, Coelho A, Diouf A, Hoctin A, Mambert M, Couffin-Cadiergues S, Deplanque D, Descamps D, Visseaux B, Desvallées M, Khan C, Diallo A, Mercier N, Paul C, Petrov-Sanchez V, Dubos F, Enouf VVE, Mouquet H, Esperou H, Jaafoura S, Papadopoulos A, Etienne M, Gigante T, Rossignol B, Guedj J, Le Nagard H, Lingas G, Neant N, Kaguelidou F, Lévy Y, Wiedemann A, Lévy Y, Wiedemann A, Levy-Marchal C, Malvy D, Noret M, Pages J, Picone O, Rossignol P, Tual C, Veislinger A, van der Werf S, Vanel N, Yazdanpanah Y, Alavoine L, Costa Y, Duval X, Ecobichon JL, Frezouls W, Ilic-Habensus E, Leclercq A, Lehacaut J, Letrou S, Mandic M, Nouroudine M, Quintin C, Rexach J, Tubiana S, Vignali V, Amat KKA, Behillil S, Enouf V, van der Werf S, Bielicki J, Bruijning P, Burdet C, Burdet C, Caumes E, Charpentier C, Damond F, Descamps D, Le Hingrat Q, Visseaux B, Coignard B, Couffin-Cadiergues S, Delmas C, Espérou H, Roufai L, Dechanet A, Houhou N, Kafif O, Kikoine J, Manchon P, Piquard V, Postolache A, Terzian Z, Lebeaux D, Lina B, Lucet JC, Malvy D, Meghadecha M, Motiejunaite J, Thy M, van Agtmael M, Bomers M, Chouchane O, Geerlings S, Goorhuis B, Grobusch MP, Harris V, Hermans SM, Hovius JW, Nellen J, Peters E, van der Poll T, Prins JM, Reijnders T, Schinkel M, Sigaloff K, Stijnis CS, van der Valk M, van Vugt M, Joost Wiersinga W, Algera AG, van Baarle F, Bos L, Botta M, de Bruin S, Bulle E, Elbers P, Fleuren L, Girbes A, Hagens L, Heunks L, Horn J, van Mourik N, Paulus F, Raasveld J, Schultz MJ, Smit M, Stilma W, Thoral P, Tsonas A, de Vries H, Bax D, Cloherty A, Beudel M, Brouwer MC, Koning R, van de Beek D, Bogaard HJ, de Brabander J, de Bree G, Bugiani M, Geerts B, Hollmann MW, Preckel B, Veelo D, Geijtenbeek T, Hafkamp F, Hamann J, Hemke R, de Jong MD, Schuurman A, Teunissen C, Vlaar APJ, Wouters D, Zwinderman AH, Abel L, Aiuti A, Muhsen S Al, Al-Mulla F, Anderson MS, Arias AA, Feldman HB, Bogunovic D, Itan Y, Bolze A, Cirulli E, Barrett KS, Washington N, Bondarenko A, Bousfiha AA, Brodin P, Bryceson Y, Bustamante CD, Butte M, Casari G, Chakravorty S, Christodoulou J, Le Mestre S, Condino-Neto A, Cooper MA, Dalgard CL, David A, DeRisi JL, DeRisi JL, Desai M, Drolet BA, Espinosa S, Fellay J, Flores C, Franco JL, Gregeren PK, Haerynck F, Hagin D, Halwani R, Heath J, Henrickson SE, Hsieh E, Imai K, Karamitros T, Kisand K, Ku CL, Lau

YL, Ling Y, Lucas CL, Maniatis T, Mansouri D, Marodi L, Meyts I, Milner J, Mironksa K, Mogensen T, Morio T, Ng LFP, Notarangelo LD, Su HC, Novelli A, Novelli G, O'Farrelly C, Okada S, Ozcelik T, de Diego RP, Planas AM, Prando C, Pujol A, Quintana-Murci L, Renia L, Renieri A, Rodríguez-Gallego C, Sancho-Shimizu V, Sankaran V, Shahrooei M, Snow A, Soler-Palacín P, Spaan AN, Tangye S, Turvey S, Uddin F, Uddin MJ, Uddin MJ, van de Beek D, Vazquez SE, Vinh DC, von Bernuth H, Zawadzki P, Casanova JL, Jing H, Tung W, Meguro K, Shaw E, Jing H, Tung W, Shafer S, Zheng L, Zhang Z, Kubo S, Chauvin SD, Meguro K, Shaw E, Lenardo M, Luthers CR, Bauman BM, Shafer S, Zheng L, Zhang Z, Kubo S, Chauvin SD, Lenardo M, Lack J, Karlins E, Hupalo DM, Rosenberger J, Sukumar G, Wilkerson MD, Zhang X. 2020. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* 370. DOI: 10.1126/science.abd4570.

Zhang Z, Zeng H, Lin J, Hu Y, Yang R, Sun J, Chen R, Chen H. 2018. Circulating LECT2 levels in newly diagnosed type 2 diabetes mellitus and their association with metabolic parameters. *Medicine (United States)* 97. DOI: 10.1097/MD.00000000000010354.

Zhao H, Lai X, Xu X, Sui K, Bu X, Ma W, Li D, Guo K, Xu J, Yao L, Li W, Su J. 2015. Histochemical analysis of testis specific gene 13 in human normal and malignant tissues. *Cell and Tissue Research* 362:653–663. DOI: 10.1007/s00441-015-2227-3.

Zhao X, Yan X, Liu Y, Zhang P, Ni X. 2016. Co-expression of mouse TMEM63A, TMEM63B and TMEM63C confers hyperosmolarity activated ion currents in HEK293 cells. *Cell Biochemistry and Function* 34:238–241. DOI: 10.1002/cbf.3185.