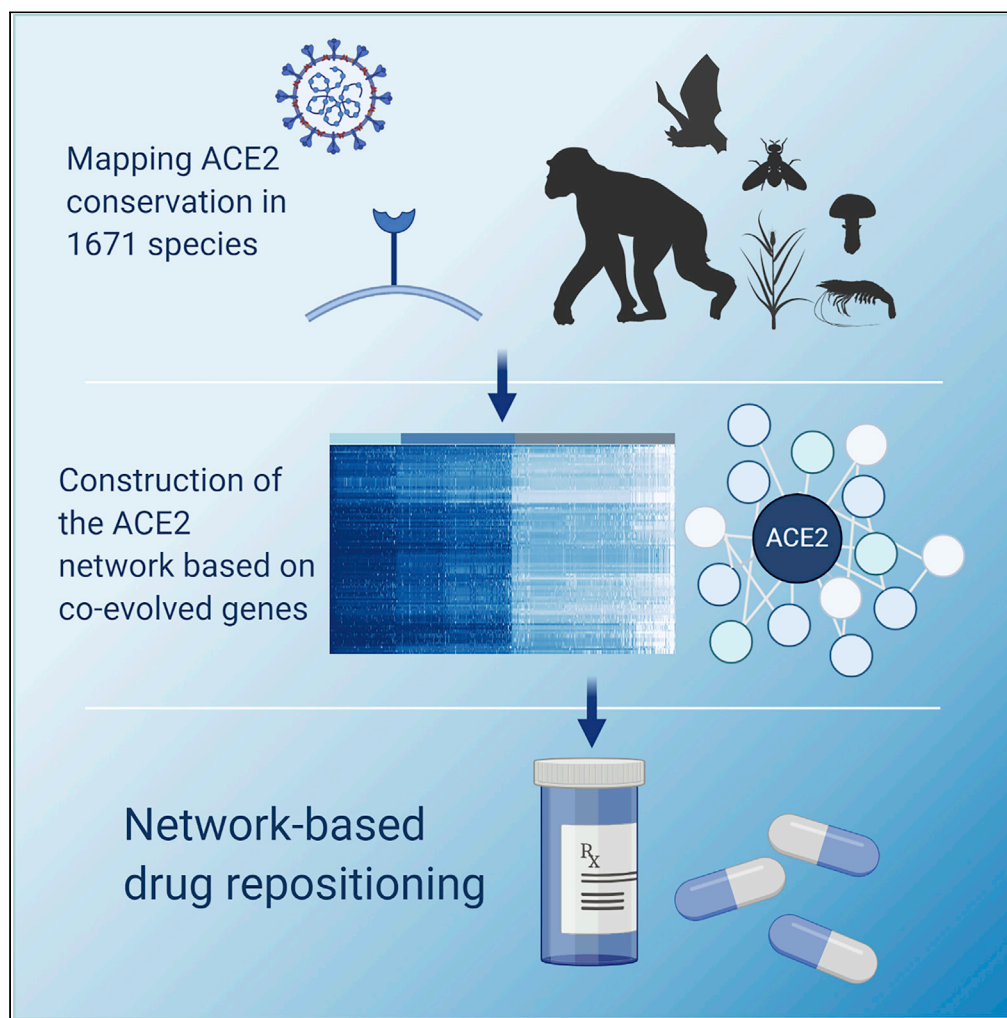


Article

ACE2 Co-evolutionary Pattern Suggests Targets for Pharmaceutical Intervention in the COVID-19 Pandemic



Maya Braun, Elad Sharon, Irene Unterman, ..., Shmuel Benenson, Alexander Vainstein, Yuval Tabach

yuvaltab@ekmd.huji.ac.il

HIGHLIGHTS

Mapping the ACE2 conservation pattern in 146 mammal reservoirs and 1,671 eukaryotes

Identification of genes that show a similar evolutionary pattern as ACE2

Generation of an ACE2 protein network using co-evolution and data integration

Drugs-to-network analyses mapped 145 drugs that perturbed the ACE2 network

Braun et al., iScience 23, 101384
August 21, 2020 © 2020 The Authors.
<https://doi.org/10.1016/j.isci.2020.101384>



Article

ACE2 Co-evolutionary Pattern Suggests Targets for Pharmaceutical Intervention in the COVID-19 Pandemic

Maya Braun,¹ Elad Sharon,^{1,3,4} Irene Unterman,^{1,4} Maya Miller,¹ Anna Mellul Shtern,¹ Shmuel Benenson,² Alexander Vainstein,³ and Yuval Tabach^{1,5,*}

SUMMARY

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spillover infection in December 2019 has caused an unprecedented pandemic. SARS-CoV-2, as other coronaviruses, binds its target cells through the angiotensin-converting enzyme 2 (ACE2) receptor. Accordingly, this makes ACE2 research essential for understanding the zoonotic nature of coronaviruses and identifying novel drugs. Here we present a systematic analysis of the ACE2 conservation and co-evolution protein network across 1,671 eukaryotes, revealing an unexpected conservation pattern in specific metazoans, plants, fungi, and protists. We identified the co-evolved protein network and pinpointed a list of drugs that target this network by using data integration from different sources. Our computational analysis found widely used drugs such as nonsteroidal anti-inflammatory drugs and vasodilators. These drugs are expected to perturb the ACE2 network affecting infectivity as well as the pathophysiology of the disease.

INTRODUCTION

Coronaviruses are a family of enveloped positive-stranded RNA viruses that have affected humans and many other mammals for over the last half century. To date, there are seven coronaviruses known to infect humans, four of which are circulating strains that are responsible for approximately 10%–30% of the cases of common cold (Paules et al., 2020). The other three strains can cause a potentially fatal disease in humans (severe acute respiratory syndrome coronavirus [SARS-CoV], Middle East respiratory syndrome coronavirus [MERS-CoV], and SARS-CoV-2). These seven strains originate from a common ancestor in bat viruses and are transferred to humans through intermediate animal hosts. SARS-CoV, the cause of a serious respiratory disease outbreak in 2002–2003, was transmitted through civet cats, whereas MERS-CoV, the cause of the 2012 outbreak, was transmitted through dromedary camels (Cui et al., 2019). Overall, more than 500 coronaviruses have been identified in bats in China, with estimates of more than 3,000 unknown bat coronavirus strains (Anthony et al., 2017).

In December 2019, a novel coronavirus was identified as the pathogen responsible for an outbreak of a severe infectious respiratory disease in Wuhan, China (Zhu et al., 2020). The virus was named SARS-CoV-2, causing the coronavirus disease COVID-19. In March 2020, the World Health Organization declared the COVID-19 outbreak as a pandemic, and immense efforts around the world are in progress to develop therapeutics. Due to the association of the earliest COVID-19 cases with the Huanan market in Wuhan, it is most plausible that the SARS-CoV-2 also emerged from an animal source. The intermediate hosts and the transmission chain in the SARS-CoV-2 spillover are yet to be elucidated. The closest SARS-like coronavirus found was sampled from the intermediate horseshoe bat (*Rhinolophus affinis*), with an ~97% identity to SARS-CoV-2. However, it displayed lesser affinity for the angiotensin-converting enzyme 2 (ACE2) receptor, identified as the key player in the SARS-CoV-2 cell entry mechanism (Hoffmann et al., 2020; Zhou et al., 2020), when compared with the coronavirus isolated from the Malayan pangolin (*Manis javanica*) found in the Guangdong province, which also displays a high homology to the human virus (Andersen et al., 2020; Zhang et al., 2020).

To date, there is no approved treatment or vaccine against SARS-CoV-2. In addition, the variability in disease severity, coupled with the fatality rates among people and in different human populations, raises the

¹Department of Developmental Biology and Cancer Research, Institute of Medical Research - Israel-Canada, The Hebrew University of Jerusalem, Jerusalem 9112102, Israel

²Department of Clinical Microbiology and Infectious Diseases, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

³Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot 76100, Israel

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: yuvaltab@ekmd.huji.ac.il
<https://doi.org/10.1016/j.isci.2020.101384>



possibility that there are genetic, environmental, nutritional, or medication usage differences (in addition to age at infection) that affect the pathophysiology of the infection. Although the complex mechanism of the disease remains an important unanswered question, substantial evidence identifies pathways of inflammation and circulation as the main drivers of morbidity (Marini and Gattinoni, 2020; Varga et al., 2020). There is an urgent need to identify other risk factors, including chronic use of certain medications, in addition to development of novel drugs. The development of new drugs is a slow, costly process, and antiviral medications are especially challenging due to the rapid evolution of virus genomes. One effective strategy is drug repurposing, which involves finding new uses for existing Food and Drug Administration (FDA)-approved drugs (Strittmatter, 2014; Pushpakom et al., 2018; Pizzorno et al., 2019). This strategy can reduce both the cost and time it would take to develop and acquire approval for COVID-19 treatments when compared with *de novo* drug discovery. Equally important, the study of known drugs might reveal those that can increase the risk for infection and worsen the outcome (Berghauer Pont et al., 2015; Thackray et al., 2018). This has extraordinary significance for drugs that are commonly used by millions of people in the world during this pandemic.

Viruses depend on their hosts to reproduce and can only survive by hijacking and rewiring the human cells' protein and gene networks. Specifically, SARS-CoV-2 binds to the cell membrane ACE2 receptor to enter human cells (Hoffmann et al., 2020; Zhou et al., 2020) and subsequently replicates inside, thus marking ACE2 as a potential target for therapeutic intervention. It may also be logically assumed that cellular pathways that were conserved during the evolution in association with the ACE2 receptor might have a role in the pathogenesis of the disease. Drugs that perturb the ACE2 network might affect the coronavirus infectivity or the severity of the diseases.

Recent work analyzed the ACE2 network using different experimental methodologies and focused on establishing the protein network (Gordon et al., 2020; Sinha et al., 2020). Although integration of this information is very beneficial, a comparative genomics approach can identify functional associations that occur across the tree of life and are not only limited to humans. Our approach to map the ACE2 network and evolution using phylogenetic profiling (PP) is independent and complementary to protein-protein interactions (PPI) (Gordon et al., 2020), or expression-based approaches (Sinha et al., 2020).

PP is an unbiased method for identifying genetic interactions through comparative genomics of thousands of organisms (Pellegrini et al., 1999). The PP of a gene describes its evolutionary conservation (in terms of sequence similarity) throughout the tree of life. If a set of genes has similar PPs, meaning their loss (sequence divergence) and retention patterns are co-dependent, they are likely to interact (Tabach et al., 2013a, 2013b; Sadreyev et al., 2015; Sherill-Rofe et al., 2019; Bloch et al., 2020).

Here we offer a novel approach for drug selection and contraindications for COVID-19. Co-evolution between proteins suggests that they are functionally coupled (Pellegrini et al., 1999). Complementary to PPI maps, perturbation of genes co-evolved with ACE2 might influence the expression, activity, or downstream target of ACE2. Consequently, viral infectivity and disease progression may be affected. Our analysis covers thousands of FDA-approved drugs to optimize the list of drugs ready for validation that can rapidly be deployed to treat the COVID-19 epidemic.

RESULTS

Characterization of ACE2 Phylogenetic Profiling and Conservation Pattern along 1,671 Eukaryotes

We mapped the ACE2 best BLASTP bit-score homologs across 1,671 eukaryotes, calculated their sequence conservation, and analyzed their protein domains compared with the human ACE2 (Figure 1) (Altschul et al., 1990). The domain analysis was conducted using the NCBI Batch Conserved Domains-Search Tool (Figures S1 and S2 and Table S1) to further assess the changes in domain structure or conservation among the species and achieve a more precise evolutionary analysis (Marchler-Bauer and Bryant, 2004).

Overall, ACE2 is conserved across the majority of the animal kingdom, with a relatively high conservation percent within vertebrates, especially mammals. As expected, the conservation rate of ACE2 decreases as the evolutionary distance increases in relation to *Homo sapiens*. This suggests the majority of the animal kingdom as potential intermediate hosts for SARS-CoV-2. Interestingly, more than 10 mammals show high divergence when compared with their clade, implying a possible protective effect from SARS-CoV-2 that

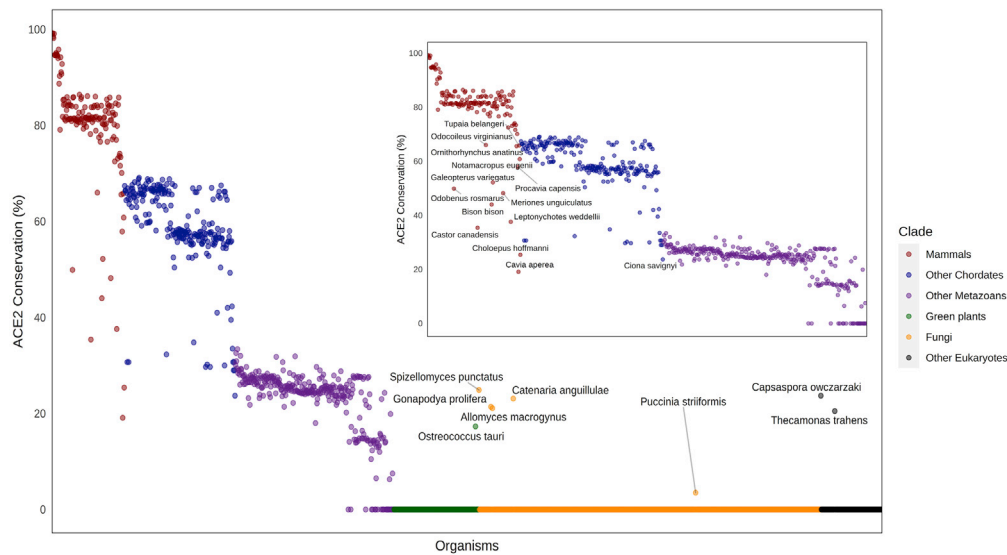


Figure 1. Evolutionary Pattern of ACE2 Proteins across 1,671 Species

The percentage of sequence similarity (y axis) to the human ACE2 protein when compared with the most similar orthologs across all eukaryotes (x axis) is presented. The Conservation score is defined as Pab/Paa , where Pab is the best BLASTP bit-score between a *Homo sapiens* protein “a” and all open reading frames of a eukaryote genome “b.” Paa is marked as the self-similarity score of the *H. sapiens* protein “a” when blasted against itself. Organisms with scores that deviated from their clade-average conservation rate are highlighted with the corresponding name. See also [Figures S1 and S2](#), and [Table S1](#).

interacts with ACE2. These species include *Odobenus rosmarus*, *Castor canadensis*, *Odocoileus virginianus*, *Bison bison*, *Galeopterus variegatus*, *Meriones unguiculatus*, *Leptonychotes weddellii*, *Ornithorhynchus anatinus*, *Procavia capensis*, *Cavia aperea*, *Tupaia belangeri*, *Notamacropus eugenii*, and *Choloepus hoffmanni*. Importantly, as far as we are aware, none of these species are known to be infected by SARS-CoV-2 through ACE2. An additional interesting observation is the lower conservation of *Ciona savignyi* within the chordates’ clade, which may also suggest divergence from the known human ACE2 function in this species. When exploring eukaryotes that further diverge from human-like fungi, plants, and different protists, there are no orthologs for ACE2. The exceptions are one green algae (*Ostreococcus tauri*), five fungi (*Spizellomyces punctatus*, *Gonapodya prolifera*, *Allomyces macrogynus*, *Catenaria anguillulae*, *Puccinia striiformis*), and two protists (*Capsaspora owczarzaki* and *Thecamonas trahens*). This observation was unexpected as ACE2 generally exists almost only in metazoans and may imply a horizontal gene transfer along the evolution. Taken together, these results point to some species for further analysis as suspected SARS-CoV-2 intermediate hosts.

Mapping Proteins that Show Similar Phylogenetic Profiling with ACE2 and Generating the ACE2 Co-evolved Network

Co-evolution across the tree of life can predict proteins that are associated with the same function, share common pathways, and contribute to corresponding diseases ([Tabach et al., 2013a, 2013b](#); [Omar et al., 2018](#); [Arkadir et al., 2019](#); [Bauer et al., 2019](#); [Sherill-Rofe et al., 2019](#)). To map the proteins that co-evolved with ACE2, we generated an extensive database of 1,671 eukaryotic proteomes and identified the set of genes that showed similar phylogenetic profiles to ACE2 ([Figures 2A and S3](#)). For each human protein, we identified the genes with the best BLASTP bit-score in each of the organisms in the database. Then we normalized the data to account for protein length and phylogenetic distance from humans (see Experimental Procedures) to generate the normalized phylogenetic profiling (NPP). We ranked all genes according to their similarity to the ACE2 profile as measured by Pearson correlation after normalization. As ACE2 is mainly conserved in vertebrates we used a clade-specific approach, which showed a higher sensitivity in detecting gene function when compared with global PP approaches ([Sherill-Rofe et al., 2019](#)). We identified ACE2 co-evolved genes in eukaryotes, metazoans, chordates, and mammals clades that contain 1,671, 688, 369, and 146 organisms, respectively ([Tables S2 and S3](#)). Finally, the top genes were further analyzed for functional enrichment.

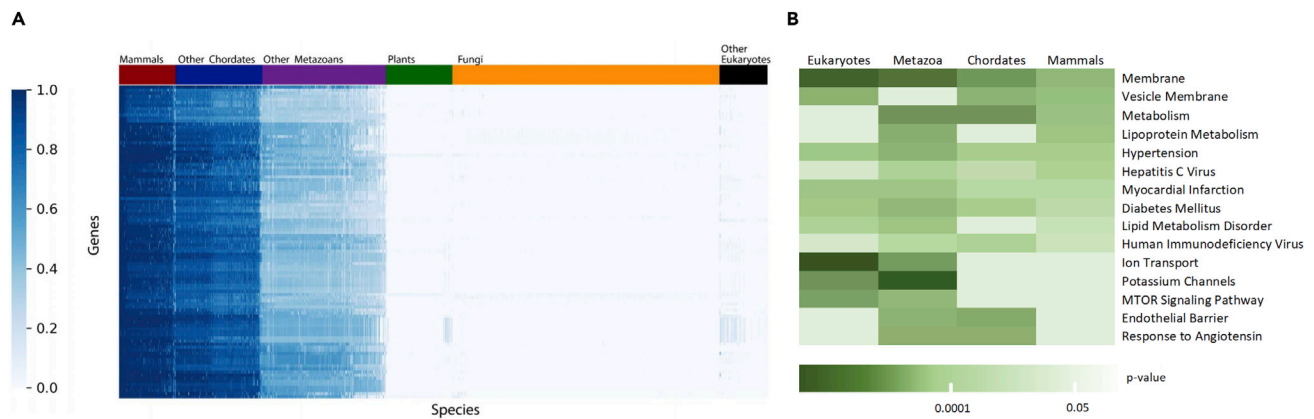


Figure 2. Evolution of the ACE2 Protein

(A) The top 100 proteins that co-evolved with ACE2, and its phylogenetic profiles across 1,671 eukaryotes. The heatmap represents the top 100 genes most correlated with ACE2, ranked by Pearson correlation of the normalized profile (log2 transformation and Z scoring). The rows are ordered based on hierarchical clustering (by the UPGMA method). Each row represents a single gene across 1,671 eukaryotes ordered by their phylogenetic distance from *Homo sapiens*. Each column represents a species. The colors indicate the relative degree of conservation from high similarity (dark blue) to not conserved (white).

(B) Enrichment analysis of the top 100 genes most correlated with ACE2 in eukaryotes, metazoans, chordates, and mammals. The top gene ontology terms and biological processes correlated in at least two clades were chosen for visualization. The color scale shows the p value cutoff levels for each biological process or disease. Deeper colors represent higher significance.

See also [Figure S3](#) and [Tables S2](#) and [S3](#).

Analysis of the top 100 most correlated genes with ACE2 in each clade revealed enrichment of common biological pathways, gene ontology (GO) biological processes, and diseases ([Figure 2B](#)) ([Ben-Ari Fuchs et al., 2016](#)). As expected, the results substantially overlapped between the four groups, stressing the similarity between them. Among the biological annotations were metabolism, membranes, endothelial barrier, ion transport, and mTOR signaling pathway. Genetic associations with the following conditions were identified as: hypertension, diabetes mellitus, myocardial infarction, and lipid metabolism disorder. All the aforementioned conditions are suspected risk factors for the severe manifestation of COVID-19 ([Chen et al., 2020](#)). These results demonstrate our ability to recapitulate the function of ACE2 and risk factors for COVID-19 infection disease using an unbiased phylogenetic profile analysis.

Identifying Drugs that Can Target the ACE2 Co-evolved Network

After establishing our ability to map disease risk factors in relation to COVID-19 pandemic, we then identified drugs that could affect the ACE2 network. For this purpose, we cross-referenced the 400 highest scoring genes (top 100 for each clade) with databases of drug-gene interactions (DGIdb, [Cotto et al., 2018](#); DrugBank, [Wishart et al., 2018](#)) to identify druggable genes. We obtained 729 optional drugs already approved by the FDA, and we further prioritized drugs according to clinical consideration (see Experimental Procedures), which resulted in 145 molecular entities ([Table S4](#)). Representative connections of the network clade-gene-drug condition are depicted in [Figure 3A](#). Analysis of the ACE2 network using STRING ([Szklarczyk et al., 2019](#)) shows that a large number of the co-evolved interactions also have supporting data for expression and proteomics in the literature ([Figure 3B](#)).

The identified 145 drugs were divided into four main categories: anti-inflammatory drugs, vasodilators, diabetes medications, and drugs targeting drug metabolism pathways, according to their clinical indication. The first group was identified with various genes playing a major role in pathways of immune response and inflammation including COX-2 (PTGS2), SLC5A8, PIK3R3, ADAMTS1, and ABCG2, which tightly co-evolved in eukaryotes and metazoans with ACE2 ([Figure 3](#)). Among the drugs targeting the aforementioned genes are the commonly used nonsteroidal anti-inflammatory drugs (NSAIDs) and inhibitors of COX-2, such as ibuprofen, aspirin, and etoricoxib. Interestingly, ADAMTS1 is targeted by statins, which are widely prescribed medications also known to have an immunomodulatory effect ([Vaughan et al., 1996](#)). In addition, ABCG2 is targeted by telaprevir and cyclosporine. The former is a direct-acting antiviral drug classified as a protease inhibitor used to treat hepatitis C ([Zeuzem et al., 2011](#)), whereas the latter inhibits viral replication of the human immunodeficiency virus and hepatitis C virus ([Xue et al., 2018](#)).

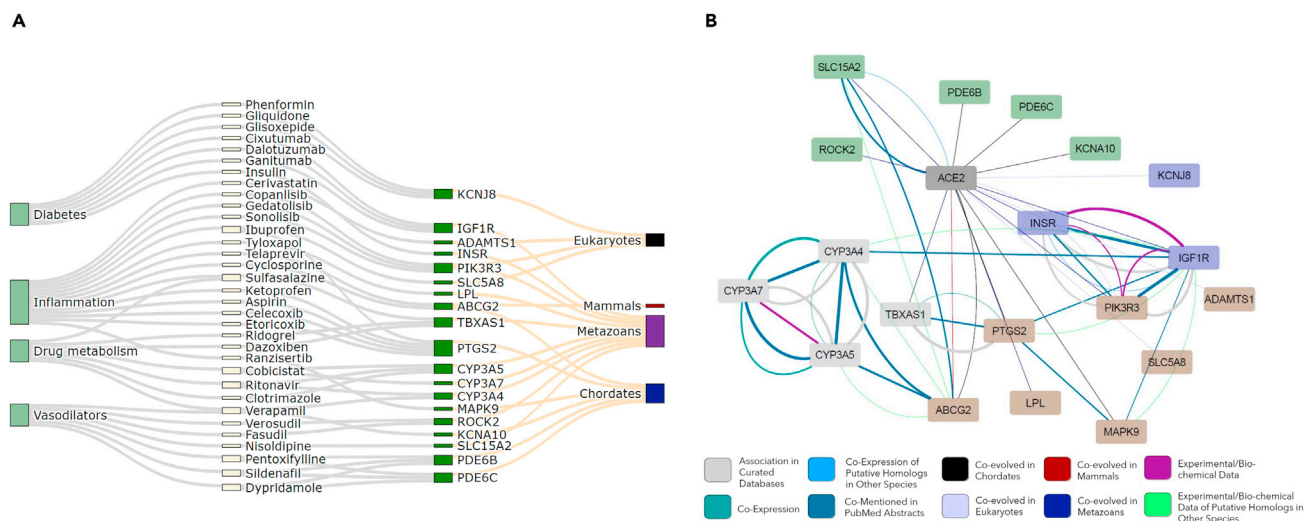


Figure 3. Representative Connections between Clade-Gene-Drug-Condition

(A) Sankey diagram of the genes that are co-evolved with ACE2 in four clades and the drugs targeting them.

(B) A subnetwork of ACE2 highlighting the genes with drug repositioning options. Node fill colors indicate subgroup affiliations inside the network: genes related to inflammation, drug metabolism, glucose metabolism, and vasodilators. Edge stroke colors represent six types of experimental evidence of the protein-protein interactions.

See also [Table S4](#).

The second appealing group of drugs in our analysis was the vasoactive drugs. Notably, a group of genes closely correlated with ACE2 in our analysis were PDE6, KCNA10, SLC15A2, and ROCK2 (Figure 3). These genes are targeted by a group of drugs with vasodilatory properties and are in widespread clinical use for hypertension, the calcium channel blockers, such as nisoldipine, verapamil, and fasudil. An additional class of drugs targeting the aforementioned genes is commonly used to treat pulmonary arterial hypertension, the phosphodiesterase (PDE) inhibitors, such as sildenafil, dypridamole, and pentoxifylline. Notably, nitric oxide (NO), a potent vasodilator, plays an important role in regulating airway function and in treating inflammatory airway diseases (Barnes, 1995) and is utilized in the treatment of COVID-19 infections.

Overall, a systematic analysis of the top 100 highest scoring genes in each clade with databases of drug-gene interactions identified 145 known drug targets of FDA-approved drugs (Table S4). The connectivity of gene-drug interactions enabled the generation of a list of drugs to target the ACE2 network with potential impacts on medical care in fighting the current evolving COVID-19 epidemic.

DISCUSSION

A major problem with coronaviruses is the repeated spillovers from the extensive animal reservoir of the virus. The conservation of ACE2 across animals probably allows high cross-species infection, which can lead to a variety of different habitats for the virus to evolve and interact with the human population. Here we present a comprehensive evolutionary analysis of ACE2. Among 1,671 organisms, we identified those that have ACE2 and the ones that are more conserved and are more likely to serve as a potential reservoir for the virus. In addition, using NPP of hundreds of species, we identified for the first time the proteins that co-evolved with ACE2. These proteins that coordinately evolve with ACE2 fit our understanding of the mechanism of action and the chronic diseases increasing the risk of COVID-19. These genes may serve as a starting point for future research to determine the modulation of ACE2 expression and function, and eventually ways to intervene pharmacologically.

PP is a robust method that identifies strong associations between genes based on their co-evolutionary course. Although the co-evolutionary signal is highly potent, PP cannot provide information about the context, type of interaction, or the directionality of the association between the genes. Future work integrating data from additional sources is needed, such as PPIs and gene expression analysis in SARS-CoV-2-infected cells with extensive validation. Nevertheless, considering the urgency to better understand

the COVID-19 pandemic, and the fact that benchwork with SARS-CoV-2 is very limited, different computational analysis are a necessity.

Using an unbiased method to investigate our comparative genomics data, we identified an association between ACE2 and various cellular networks. Within these networks we found druggable targets that may influence susceptibility to poor outcomes of viral infection. Among the drugs are COX inhibitors and vasodilators, medications that might be particularly relevant as the severe course of COVID-19-related disease is characterized by uncontrolled inflammation and circulatory collapse. Indeed, with almost 10 million infected patients in the world, and extensive descriptions of the clinical course and the pathologic changes in the disease, it became more evident that endothelial damage and thromboembolic complications are involved in the severe course of the disease. The vasomotor and thromboembolic pathogenesis of severe COVID-19 probably accounts for the prevalent risk factors of obesity, diabetes, and hypertension, all of them being associated with endothelial dysfunction. These pathways were correctly identified by our analysis before the above data became available. These pathways are targeted by aspirin, statins, and vasodilators. The aforementioned classes of drugs are in extensive clinical use to treat cardiovascular and inflammatory conditions (e.g., over 30 billion doses of NSAIDs are consumed annually in the United States alone).

Awareness of the possible association between these medications and the pathways involved in COVID-19 infection might lead to significant clinical impacts to maximize their positive effect. In addition, an antiviral medication was found to be linked to the network and might be applied to the treatment of COVID-19, subject to appropriate clinical testing.

In conclusion, this is the first study using PP to identify ACE2 conservation and its co-evolved proteins across millions of years of evolution. Many of the drugs identified show a clear association with the disease, and their effect on patients should be examined. Overall, due to the urgent need for a comprehensive approach to treat COVID-19 during the current epidemic, we believe that our work provides a list of existing drugs to be evaluated. Nevertheless, additional data are needed to prove these observations.

Limitations of the Study

Computational and high-throughput approaches including PP analysis can provide a hint and suggest a mechanism. However, these observations, even when biologically sound, still may not be clinically relevant, and further extensive work is required.

Resource Availability

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Yuval Tabach (yuvaltab@ekmd.huji.ac.il).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Original data have been deposited to Mendeley Data: <http://doi.org/10.17632/zrzh6x74th.4>.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101384>.

ACKNOWLEDGMENTS

Funding was received from the Israel Science Foundation (grant agreement 1591/19), and the Israel Innovation Authority under the R&D Plans of Industrial Products for the Prevention and Treatment of the COVID-19 (grant agreement 70273). The graphical abstract was created by BioRender.com.

AUTHOR CONTRIBUTIONS

Conceptualization, M.B. and Y.T.; Methodology, M.B., E.S., I.U., and Y.T.; Software, E.S. and I.U.; Formal Analysis, M.B.; Investigation, M.B., M.M., E.S., and I.U.; Resources, A.M.S. and A.V. Writing – Original Draft, M.B., E.S., and M.M.; Writing – Review & Editing, M.B., M.M., S.B., and Y.T.; Visualization, I.U., E.S., M.B., and M.M.; Supervision, Y.T.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 3, 2020

Revised: June 26, 2020

Accepted: July 16, 2020

Published: August 21, 2020

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Andersen, K.G., Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* *89*, 44–48.
- Anthony, S.J., Johnson, C.K., Greig, D.J., Kramer, S., Che, X., Wells, H., Hicks, A.L., Joly, D.O., Wolfe, N.D., Daszak, P., et al. (2017). Global patterns in coronavirus diversity. *Virus Evol.* *3*, 1–15.
- Arkadir, D., Lossos, A., Rahat, D., Abu Snineh, M., Schueler-Furman, O., Nitschke, S., Minassian, B.A., Sadaka, Y., Lerer, I., Tabach, Y., and Meiner, V. (2019). MYORG is associated with recessive primary familial brain calcification. *Ann. Clin. Transl. Neurol.* <https://doi.org/10.1002/acn3.684>.
- Barnes, P.J. (1995). Nitric oxide and airway disease. *Ann. Med.* <https://doi.org/10.3109/07853899509002592>.
- Bauer, M., Rahat, D., Zisman, E., Tabach, Y., Lossos, A., Meiner, V., and Arkadir, D. (2019). MYORG mutations: a major cause of recessive primary familial brain calcification. *Curr. Neurol. Neurosci. Rep.* <https://doi.org/10.1007/s11910-019-0986-z>.
- Ben-Ari Fuchs, S., Lieder, I., Stelzer, G., Mazor, Y., Buzhor, E., Kaplan, S., Bogoch, Y., Plaschkes, I., Shitrit, A., Rappaport, N., et al. (2016). GeneAnalytics: an integrative gene set analysis tool for next generation sequencing, RNAseq and microarray data. *OMICS.* <https://doi.org/10.1089/omi.2015.0168>.
- Berghauer Pont, L.M., Balvers, R.K., Kloezeman, J.J., Nowicki, M.O., van den Bossche, W., Kremer, A., Wakimoto, H., van den Hoogen, B.G., Leenstra, S., Dirven, C.M., et al. (2015). In vitro screening of clinical drugs identifies sensitizers of oncolytic viral therapy in glioblastoma stem-like cells. *Gene Ther.* <https://doi.org/10.1038/gt.2015.72>.
- Bloch, I., Sherill-Rofe, D., Stupp, D., Unterman, I., Beer, H., Sharon, E., and Tabach, Y. (2020). Optimization of Co-evolution analysis through phylogenetic profiling reveals pathway-specific signals. *Bioinformatics.* <https://doi.org/10.1093/bioinformatics/btaa281>.
- Chen, T., Wu, D., Chen, H., Yan, W., Yang, D., Chen, G., Ma, K., Xu, D., Yu, H., Wang, H., et al. (2020). Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ.* <https://doi.org/10.1136/bmj.m1091>.
- Cotto, K.C., Wagner, A.H., Feng, Y.Y., Kiwala, S., Coffman, A.C., Spies, G., Wollam, A., Spies, N.C., Griffith, O.L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkx1143>.
- Cui, J., Li, F., and Shi, Z.L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* <https://doi.org/10.1038/s41579-018-0118-9>.
- Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature.* <https://doi.org/10.1038/s41586-020-2286-9>.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically-proven protease inhibitor. *Cell.* 1–10, <https://doi.org/10.1016/j.cell.2020.02.052>.
- Marchler-Bauer, A., and Bryant, S.H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkh454>.
- Marini, J.J., and Gattinoni, L. (2020). Management of COVID-19 respiratory distress. *JAMA.* <https://doi.org/10.1001/jama.2020.6825>.
- Omar, I., Guterman-Ram, G., Rahat, D., Tabach, Y., Berger, M., and Levaot, N. (2018). Schlafen2 mutation in mice causes an osteopetrotic phenotype due to a decrease in the number of osteoclast progenitors. *Sci. Rep.* <https://doi.org/10.1038/s41598-018-31428-z>.
- Paules, C.I., Marston, H.D., and Fauci, A.S. (2020). Coronavirus infections—more than just the common cold. *JAMA.* <https://doi.org/10.1001/jama.2020.0757>.
- Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., and Yeates, T.O. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U.S.A.* <https://doi.org/10.1073/pnas.96.8.4285>.
- Pizzorno, A., Padey, B., Terrier, O., and Rosa-Calatrava, M. (2019). Drug repurposing approaches for the treatment of influenza viral infection: reviving old drugs to fight against a long-lived enemy. *Front. Immunol.* <https://doi.org/10.3389/fimmu.2019.00531>.
- Pushpakom, S., Iorio, F., Eyers, P.A., Escott, K.J., Hopper, S., Wells, A., Doig, A., Williams, T., Latimer, J., McNamee, C., et al. (2018). Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* <https://doi.org/10.1038/nrd.2018.168>.
- Sadreyev, I.R., Ji, F., Cohen, E., Ruvkun, G., and Tabach, Y. (2015). PhyloGene server for identification and visualization of co-evolving proteins using normalized phylogenetic profiles. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkv452>.
- Sherill-Rofe, D., Sherill-Rofe, D., Rahat, D., Findlay, S., Mellul, A., Guberman, I., Braun, M., Bloch, I., Lalezari, A., Samiei, A., et al. (2019). Mapping global and local coevolution across 600 species to identify novel homologous recombination repair genes. *Genome Res.* <https://doi.org/10.1101/gr.241414.118>.
- Sinha, S., Cheng, K., Aldape, K., Schiff, E., and Ruppin, E. (2020). Systematic cell line-based identification of drugs modifying ACE2 expression. Preprints. <https://doi.org/10.20944/PREPRINTS202003.0446.V1>.
- Strittmatter, S.M. (2014). Overcoming drug development bottlenecks with repurposing: old drugs learn new tricks. *Nat. Med.* <https://doi.org/10.1038/nm.3595>.
- Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide

experimental datasets. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gky1131>.

Tabach, Y., Golan, T., Hernández-Hernández, A., Messer, A.R., Fukuda, T., Kouznetsova, A., Liu, J.G., Lilienthal, I., Levy, C., and Ruvkun, G. (2013a). Human disease locus discovery and mapping to molecular pathways through phylogenetic profiling. *Mol. Syst. Biol.* *9*, 1–17.

Tabach, Y., Billi, A.C., Hayes, G.D., Newman, M.A., Zuk, O., Gabel, H., Kamath, R., Yacoby, K., Chapman, B., Garcia, S.M., et al. (2013b). Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature*. <https://doi.org/10.1038/nature11779>.

Thackray, L.B., Handley, S.A., Gorman, M.J., Poddar, S., Bagadia, P., Briseño, C.G., Theisen, D.J., Tan, Q., Hykes, B.L., Jr., Lin, H., et al. (2018). Oral antibiotic treatment of mice exacerbates the disease severity of multiple flavivirus infections. *Cell Rep.* <https://doi.org/10.1016/j.celrep.2018.03.001>.

Varga, Z., Flammer, A.J., Steiger, P., Haberecker, M., Andermatt, R., Zinkernagel, A.S., Mehra, M.R., Schuepbach, R.A., Ruschitzka, F., and Moch, H. (2020). Endothelial cell infection and endotheliitis in COVID-19. *Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)30937-5](https://doi.org/10.1016/S0140-6736(20)30937-5).

Vaughan, C.J., Murphy, M.B., and Buckley, B.M. (1996). Statins do more than just lower cholesterol. *Lancet*. [https://doi.org/10.1016/S0140-6736\(96\)05190-2](https://doi.org/10.1016/S0140-6736(96)05190-2).

Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkx1037>.

Xue, C., Sowden, M.P., and Berk, B.C. (2018). Extracellular and intracellular cyclophilin A, native and post-translationally modified, show diverse and specific pathological roles in diseases. *Arterioscler. Thromb. Vasc. Biol.* <https://doi.org/10.1161/ATVBAHA.117.310661>.

Zeuzem, S., Andreone, P., Pol, S., Lawitz, E., Diago, M., Roberts, S., Focaccia, R., Younossi, Z., Foster, G.R., Horban, A., et al. (2011). Telaprevir for retreatment of HCV infection. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa1013086>.

Zhang, T., Wu, Q., and Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Curr. Biol.* 1–6, <https://doi.org/10.1016/j.cub.2020.03.022>.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. <https://doi.org/10.1038/s41586-020-2012-7>.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* <https://doi.org/10.1056/nejmoa2001017>.

iScience, Volume 23

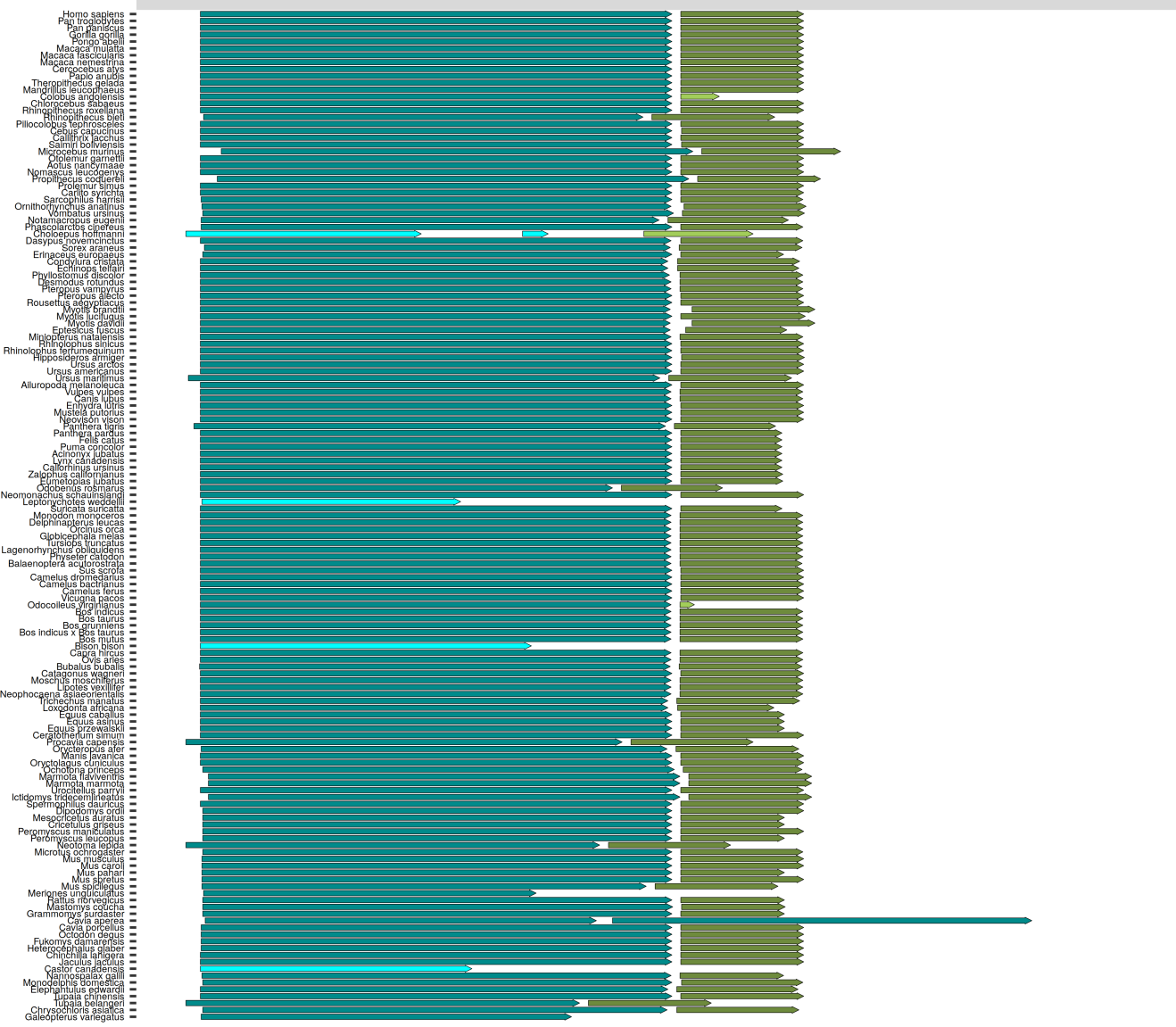
Supplemental Information

**ACE2 Co-evolutionary Pattern Suggests
Targets for Pharmaceutical Intervention
in the COVID-19 Pandemic**

Maya Braun, Elad Sharon, Irene Unterman, Maya Miller, Anna Mellul Shtern, Shmuel Benenson, Alexander Vainstein, and Yuval Tabach

Figure S1. Domain analysis of *Homo sapiens* ACE2 best BLASTP bit-score across mammals and chordates, related to Figure 1. The sequences were analysed using the NCBI Batch Conserved Domains-Search Tool.

Mammals



Other Chordates

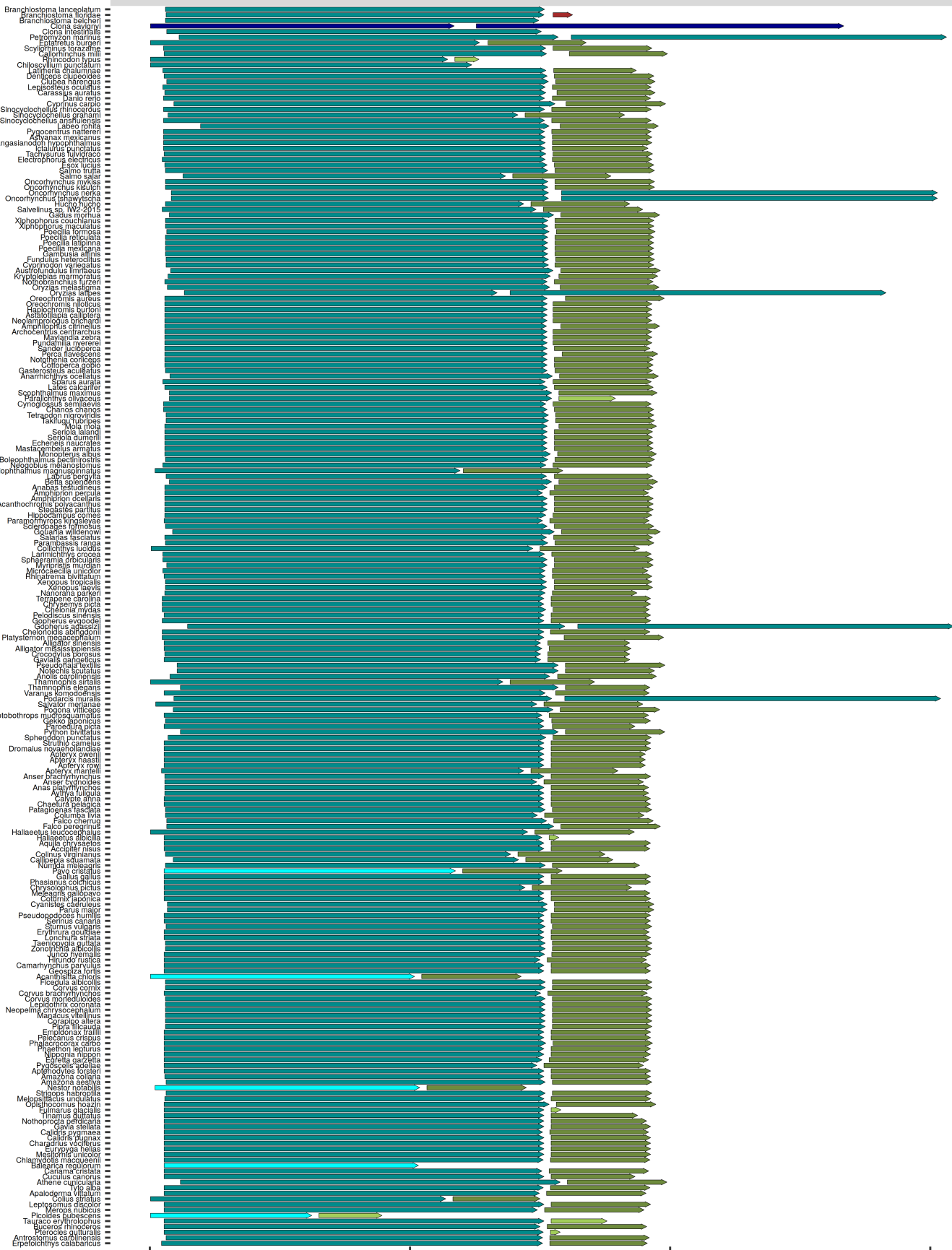
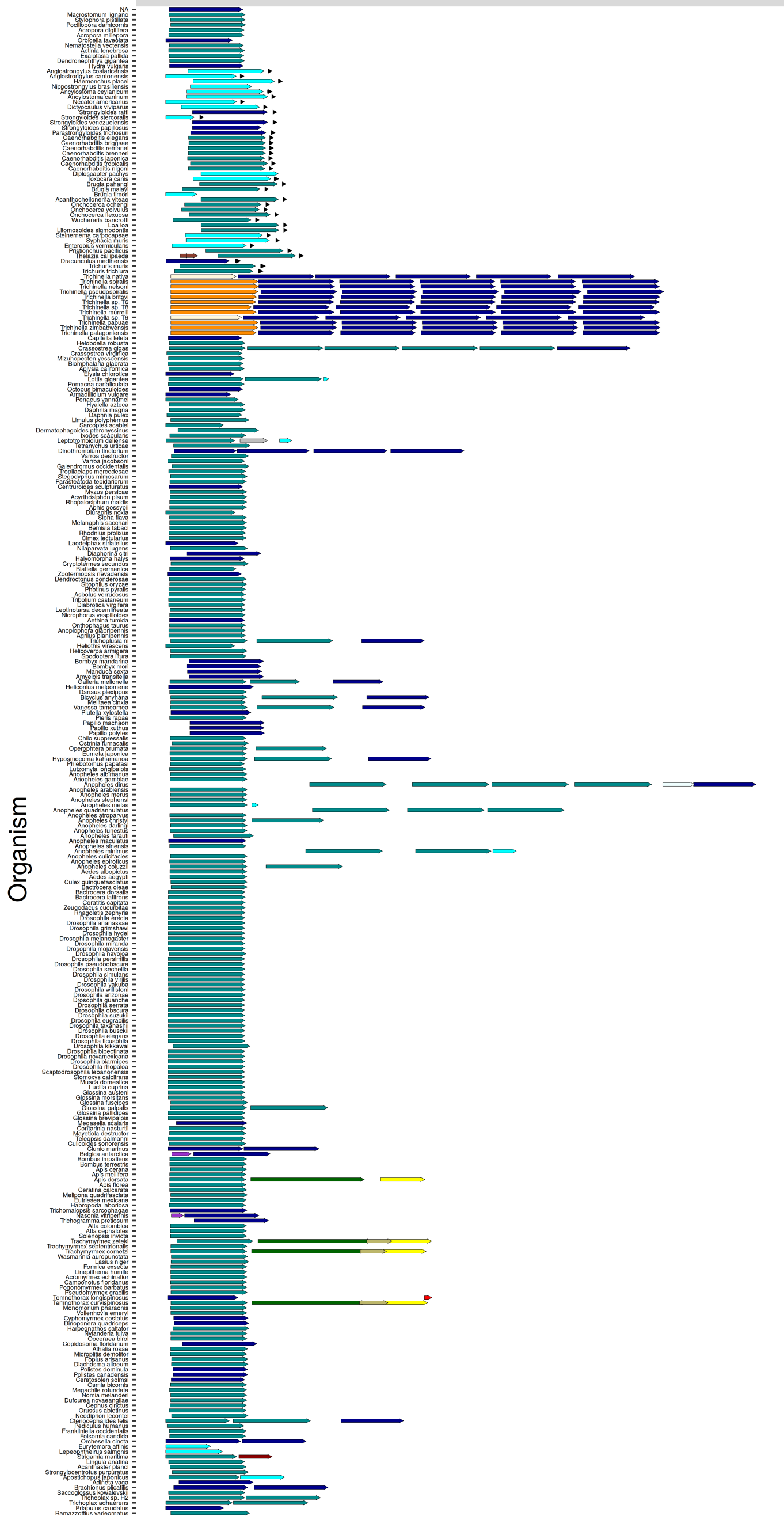
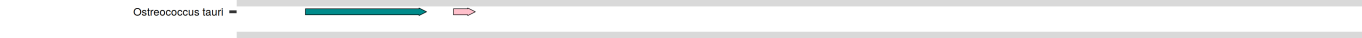


Figure S2. Domain analysis of *Homo sapiens* ACE2 best BLASTP bit-score across metazoans and selected species with divergent results among plants, fungi, and other eukaryotes, related to Figure 1. The sequences were analysed using the NCBI Batch Conserved Domains-Search Tool.

Other Metazoans



Green plants



Fungi



Other Eukaryotes



Figure S3. The top 100 proteins that co-evolved with ACE2 and its phylogenetic profiles across (A) 688 metazoans, (B) 369 chordates, and (C) 146 mammals, related to Figure 2. The heatmap represents the top 100 genes most correlated with ACE2, ranked by Pearson correlation of the normalized profile (log2 transformation and z-scoring). The rows are ordered based on hierarchical clustering (by the UPGMA method).

TRANSPARENT METHODS

Establishing the genomic database of 1671 eukaryotes

To collect the extensive details of all the organism proteomes, we integrated information from three different databases: Ensembl (Yates *et al.*, 2020), NCBI Genomes Refseq (Sayers *et al.*, 2020) and Uniprot “reference proteomes” (Chen *et al.*, 2011). We further integrated the pre-formatted BLAST NR database (accessed 19 February 2019) using the script “update_blastdb.pl” from BLAST 2.7.1+ version (Camacho *et al.*, 2009). The inferred proteomes were merged to the species level and each organism was filtered from duplicated sequences, resulting in 1671 organisms. In order to extract specific taxonomies from the compressed pre-formatted BLAST NR database, we used the “taxonkit” tool (Shen and Xiong, 2019) that maps the sequence identifiers to the relevant species and extracts the relevant fasta files. The described above procedure eliminates biases that exist in a species database and prevents false annotation of protein loss.

Establishing the conservation pattern (phylogenetic profiles) of ACE2 and the other human genes

A BLASTP (Altschul *et al.*, 1990) comparison generates a matrix P of size $20,294 \times 1671$, where each entry P_{ab} is the best BLASTP bit score between a human protein sequence ‘a’ and the top result in organism ‘b’. The BLASTP scores provide a continuous phylogenetic profile, indicating homology level at each species. This approach is more sensitive than traditional binary phylogenetic profiles, which are based only on a comparison of the presence or absence pattern of suites of factors in specific clades of organisms. To reduce the influence of random matches in the phylogenetic profiles, low BLASTP bit scores (E-value $> 10^{-5}$) were assigned a value of 0 ($P_{ab} = 0$).

1. *Normalizing the BLASTP bit scores for protein length:* Since the BLASTP score depends linearly on the length of protein ‘a’, long alignments would tend to have higher scores independent of whether the aligned segments show sequence similarity, resulting in a bias towards longer proteins. We therefore normalized the phylogenetic profile matrix values to remove biases resulting from variations in protein lengths. In addition to P_{ab} , the best BLASTP bit score between a *Homo Sapiens* protein ‘a’ and all ORFs of a eukaryote genome ‘b’, we computed P_{aa} , defined as the self-similarity score of the *Homo Sapiens* protein ‘a’ when blasted against itself. In Length Normalized Phylogenetic Profiling (LNPP), the normalized phylogenetic profile matrix, each entry in the row corresponding to protein ‘a’ is computed as $LNPP_{ab} = P_{ab}/P_{aa}$. The normalized BLASTP score represents the ratio of the observed BLASTP score and the best possible BLASTP score of the same length (the self-similarity score).

2. *Normalizing for organisms with different evolutionary distance:* A second normalization procedure was applied in order to compensate for the different protein similarity (i.e. score) expected when human proteins are compared to proteins from eukaryotes of highly variable evolutionary distance. For this purpose, we normalized the log values in each column ‘b’ (i.e. each organism) by subtracting their average μ_b and dividing by their standard deviation σ_b , yielding: $NPP_{ab} = (\log_2(LNPP_{ab} + 1) - \mu_b) / \sigma_b$

Co-evolution analysis

The species taxonomy was ordered in a descending evolutionary distance from *homo sapiens* based on the NCBI Common Taxonomy Tree architecture. To identify druggable gene targets

that are coevolved with ACE2 in the context of a specific clade, we analyzed four subgroups of the complete NPP matrix corresponding to the following clades: mammals, chordates, metazoans, and all eukaryotes, including 146, 369, 688, and 1671 organisms, respectively. In each clade tested, we ranked all human genes by the similarity of their PP to ACE2. Pearson correlation was used as the NPP metric. Known interactions to ACE2 were mapped with STRING (Szklarczyk *et al.*, 2019), and plotted using Cytoscape (Shannon *et al.*, 2003).

Enrichment analysis

The top 100 genes most correlated with ACE2 in each clade were analyzed independently for enrichment using GeneAnalytics (Fuchs *et al.*, 2016). Gene set - disease relationship analysis was based on the information available in MalaCards (Rappaport *et al.*, 2017), a human disease database integrating data from over 60 sources. Pathway relationships were analyzed using PathCards (Belinky *et al.*, 2015), a gene-content algorithm unifying thousands of pathways clustered into SuperPaths. Gene ontology (GO) terms available in the GO project and integrated in GeneCards (Stelzer *et al.*, 2016) were analyzed for enrichment. The predominant common pathways, GO terms, and disease relationships were chosen for presentation.

Drug-target network

The 100 highest scoring genes in each clade were analyzed to identify drug targets using DrugBank (Wishart *et al.*, 2018) and the drug-gene interaction database (DGIdb) (Cotto *et al.*, 2018). We collected the drug-target information from these databases, integrating over 40,000 genes and 13,000 drug entries, and cross-referenced it with the formed ACE2 network. We filtered the results for FDA approved drugs and drugs with identified gene-drug interactions, resulting in a list of 729 drugs. The final list of 145 drugs was generated based on manual ranking of drugs according to the following criteria: 1) Correlation to pathways related to the ACE2 receptor function, 2) Involvement in pathways related to ACE2 synthesis, 3) Clinical association to drugs treating pulmonary vascular disorders, and 4) Safety and efficacy profiles of drugs when used for conventional indications.

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*. doi: 10.1016/S0022-2836(05)80360-2.

Belinky, F. *et al.* (2015) 'PathCards: Multi-source consolidation of human biological pathways', *Database*. doi: 10.1093/database/bav006.

Camacho, C. *et al.* (2009) 'BLAST+: Architecture and applications', *BMC Bioinformatics*. doi: 10.1186/1471-2105-10-421.

Chen, C. *et al.* (2011) 'Representative Proteomes: A Stable, Scalable and Unbiased proteome set for sequence analysis and functional annotation', *PLoS ONE*. doi: 10.1371/journal.pone.0018910.

Cotto, K. C. *et al.* (2018) 'DGIdb 3.0: A redesign and expansion of the drug-gene interaction database', *Nucleic Acids Research*. doi: 10.1093/nar/gkx1143.

Fuchs, S. B. A. *et al.* (2016) 'GeneAnalytics: An Integrative Gene Set Analysis Tool for Next Generation Sequencing, RNAseq and Microarray Data', *OMICS A Journal of Integrative Biology*. doi: 10.1089/omi.2015.0168.

Rappaport, N. *et al.* (2017) 'MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search', *Nucleic Acids Research*. doi: 10.1093/nar/gkw1012.

Sayers, E. W. *et al.* (2020) 'Database resources of the National Center for Biotechnology

Information', *Nucleic acids research*. doi: 10.1093/nar/gkz899.

Shannon, P. *et al.* (2003) 'Cytoscape: A software Environment for integrated models of biomolecular interaction networks', *Genome Research*. doi: 10.1101/gr.1239303.

Shen, W. and Xiong, J. (2019) 'TaxonKit: a cross-platform and efficient NCBI taxonomy toolkit', *bioRxiv*. doi: 10.1101/513523.

Stelzer, G. *et al.* (2016) 'The GeneCards suite: From gene data mining to disease genome sequence analyses', *Current Protocols in Bioinformatics*. doi: 10.1002/cpbi.5.

Szklarczyk, D. *et al.* (2019) 'STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic Acids Research*. doi: 10.1093/nar/gky1131.

Wishart, D. S. *et al.* (2018) 'DrugBank 5.0: A major update to the DrugBank database for 2018', *Nucleic Acids Research*. doi: 10.1093/nar/gkx1037.

Yates, A. D. *et al.* (2020) 'Ensembl 2020', *Nucleic acids research*. doi: 10.1093/nar/gkz966.