Identification of evolutionarily stable functional and immunogenic sites across the SARS-CoV-2 proteome and the greater coronavirus family

Chen Wang
Baylor College of Medicine https://orcid.org/0000-0001-5769-2077
Daniel M. Konecki
Baylor College of Medicine https://orcid.org/0000-0002-9729-5217
David C. Marciano (david.marciano@bcm.edu)
Baylor College of Medicine https://orcid.org/0000-0001-5237-5144
Harikumar Govindarajan
Baylor College of Medicine https://orcid.org/0000-0001-6075-5884
Amanda M. Williams
Baylor College of Medicine https://orcid.org/0000-0002-9212-5980
Brigitta Wastuwidyaningtyas
Baylor College of Medicine https://orcid.org/0000-0001-7270-1891
Thomas Bourquard
Baylor College of Medicine https://orcid.org/0000-0002-9670-711X
Panagiotis Katsonis
Baylor College of Medicine https://orcid.org/0000-0002-7172-1644
Olivier Lichtarge (lichtarge@bcm.edu)
Baylor College of Medicine https://orcid.org/0000-0003-4057-7122

Research Article

Keywords: SARS-CoV-2, COVID-19, coronavirus, epitopes, sequence analysis

Posted Date: February 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-95030/v3

Identification of evolutionarily stable functional and immunogenic sites across the SARS-CoV-2 proteome and the greater coronavirus family

Chen Wang[a,1], Daniel M. Konecki[b,1], David C. Marciano[a,1,*], Harikumar Govindarajan[a], Amanda M. Williams[c], Brigitta Wastuwidyaningtyas[a], Thomas Bourquard[a,d], Panagiotis Katsonis[a] and Olivier Lichtarge[a,b,c,e,*]

[a]Department of Molecular and Human Genetics, [b]Quantitative and Computational Biosciences Graduate Program, and [c]Cancer and Cell Biology Graduate Program, Baylor College of Medicine, Houston, TX 77030, USA; [d]MAbSilico, Nouzilly, Centre, France, EU; [e]Computational and Integrative Biomedical Research Center, Baylor College of Medicine, Houston, TX 77030, USA

[1]These authors contributed equally

*Correspondence: david.marciano@bcm.edu (D.C.M), lichtarge@bcm.edu (O.L.)

Chen Wang (0000-0001-5769-2077), Daniel Konecki (0000-0002-9729-5217), David Marciano (0000-0001-5237-5144), Harikumar Govindarajan (0000-0001-6075-5884), Amanda Williams (0000-0002-9212-5980), Brigitta Wastuwidyaningtyas (0000-0001-7270-1891), Thomas Bourquard (0000-0002-9670-711X), Panagiotis Katsonis (0000-0002-7172-1644), Olivier Lichtarge (0000-0003-4057-7122)

Keywords

SARS-CoV-2, COVID-19, coronavirus, epitopes, sequence analysis

Author Contributions

C.W., D.M.K., D.C.M., and O.L. designed research; C.W., D.M.K., and D.C.M. performed research; C.W., D.M.K., H.G., T.B., and P.K. contributed new analytic tools; C.W., D.M.K., D.C.M., H.G., T.B., P.K., and O.L analyzed data; and C.W., D.M.K., D.C.M., A.M.W., B.W., and O.L. wrote the paper.

This file includes:

Main Text
Figures 1 to 4

Other supplementary materials for this manuscript include the following:

Supplementary text with Supporting materials and methods, Figures S1 to S8 and Legends for Datasets S1 to S9
Datasets S1 to S9

Abstract

Since the first recognized case of COVID-19, more than 100 million people have been infected worldwide. Global efforts in drug and vaccine development to fight the disease have yielded vaccines and drug candidates to cure COVID-19. However, the spread of SARS-CoV-2 variants threatens the continued efficacy of these treatments. In order to address this, we interrogate the evolutionary history of the entire SARS-CoV-2 proteome to identify evolutionarily conserved functional sites that can inform the search for treatments with broader coverage across the coronavirus family. Combining this information with the mutations observed in the current COVID-19 outbreak, we systematically and comprehensively define evolutionarily stable sites that may provide useful drug and vaccine targets and which are less likely to be compromised by the emergence of new virus strains. Several experimentally-validated effective drugs interact with these proposed target sites. In addition, the same evolutionary information can prioritize cross reactive antigens that are useful in directing multi-epitope vaccine strategies to illicit broadly neutralizing immune responses to the betacoronavirus family. Although the results are focused on SARS-CoV-2, these approaches stem from evolutionary principles that are agnostic to the organism or infective agent.

Significance Statement

By examining past evolutionary pressures in the coronavirus family and in the present SARS-CoV-2 outbreak, we identified functional sites in the SARS-CoV-2 proteome that are not affected by variants in the current outbreak. These sites can be targeted for small molecule docking, used

for pan-coronavirus/betacoronavirus vaccine and monoclonal antibody (mAb) development, provide templates for mimetic peptides, and offer genetic targets to generate attenuated virus.

Main Text

INTRODUCTION

COVID-19 is a worldwide affliction. Since first being reported in December 2019 in Wuhan, Hubei province, China, the World Health Organization (WHO) has tallied more than 2 million COVID-19 related deaths and over 100 million infections worldwide (as of February 8th, 2021) (Dong et al., 2020). Although timely public health interventions can successfully curtail incidence, the threat of subsequent waves of infections remains widespread and the emergence of new strains that evade current treatments is becoming a likely scenario (Dong et al., 2020; Kraemer et al., 2020; Wibmer et al., 2021; Wu et al., 2021; J. Zhang et al., 2020). The novel betacoronavirus (SARS-CoV-2) that is causing the pandemic is closely related to other known human coronavirus pathogens SARS-CoV, MERS-CoV (Chan et al., 2020; Lu et al., 2020), HCoV OC43, HKU1 and is more distantly related to the human infectious alphacoronaviruses HCoV 229E and HCoV NL63 (Lei et al., 2018). Finding ways to control and prevent further infection are top priorities which include the targeted discovery of drugs that impair viral mechanisms (Youngchang Kim et al., 2020; Li et al., 2020; Rut et al., 2020) and antigenic epitopes through which vaccines raise immunity (Mullard, 2020; Poh et al., 2020; van Doremalen et al., 2020). This study addresses both by utilizing evolutionary information from SARS-CoV-2 sequence and structural data to search for actionable functional sites for each protein in the SARS-CoV-2 genome.

In a first application, we note that the approval of new drugs under normal circumstances often takes more than 10 years (Dhama et al., 2020; Pillaiyar et al., 2020). In order to hasten the response, many current clinical trials for COVID-19 enlist antiviral agents that have targeted Zika, SARS-CoV, Ebola, and MERS-CoV in the past (Dhama et al., 2020; Jogalekar et al., 2020). In order to test more varieties of potential drugs, some studies screened thousands of clinical-stage or FDA-approved small molecules for antiviral activity, hoping to repurpose some of the top hits for COVID-19 treatment (Riva et al., 2020; White et al., 2021). However, the antiviral activity in these large-scale screens may, in part, be cell-line specific (Hoffmann et al., 2020), and therefore of unclear clinical relevance. Another approach to screen potential drugs for repurposing is to perform docking (Goodsell et al., 2020) of clinical-stage or FDA-approved drugs to the SARS-CoV-2 proteome (S. Gupta et al., 2020; Ortega et al., 2020). However, selection of the correct binding sites on the target proteins is crucial and difficult as protein surface cavities far exceed actual ligand binding sites that modulate function (Gupta et al., 2018). Here we systematically suggest potential drug target sites for most SARS-CoV-2 proteins based on evolutionary information. As these sites are chosen for their conserved functional roles, broad pan-coronavirus/betacoronavirus relevance, and minimal variability across all known SARS-CoV-2 variants, they should be prioritized in docking studies for drug repurposing.

In a second application, we note that understanding the immune response to SARS-CoV-2 infection is critical for vaccine development (Grifoni et al., 2020b). Most early SARS-CoV-2 immune epitope discovery studies rely heavily on bioinformatic prediction tools as well as sequence and epitope work already done in SARS-CoV and MERS-CoV. B-cell linear and discontinuous epitope prediction tools have been used by researchers to identify possible SARS-CoV-2 epitopes (Ahmed et al., 2020; Bhattacharya et al., 2020; Grifoni et al., 2020a). Several more recent studies experimentally determined SARS-CoV-2 immune epitopes (Le Bert et al., 2020; Nolan et al., 2020; Poh et al., 2020). Interestingly, several groups have reported significant T-cell reactivity against SARS-CoV-2 epitopes in individuals without virus exposure (Grifoni et al., 2020b; Le Bert et al.,

2020; Mateus et al., 2020; Meckiff et al., 2020). Mateus et al. suggested that this could be due to cross reactivity between SARS-CoV-2 and other common human coronaviruses, such as OC43, HKU1, NL63 and 229E (Mateus et al., 2020). Here we report an evolutionary metric, which can accurately separate cross-reactive epitopes from those that are not, and use this metric to suggest potential cross-reactive epitopes in SARS-CoV-2. Prioritizing these cross-reactive epitopes in vaccine development can potentially lead to broadly neutralizing immunity across the betacoronavirus family, provide starting points for mAb development and offer guidance in updating current vaccines in the face of variants which make them less effective. The AstraZeneca, Moderna, and Pfizer vaccines currently being administered target the SARS-CoV-2 Spike glycoprotein (Baden et al., 2021; Polack et al., 2020; Voysey et al., 2021), while the cross-reactive epitopes highlighted here occur in less variable regions of the Spike protein and in other SARS-CoV-2 proteins.

Here, we use the Evolutionary Trace (ET) method, which predicts the importance of protein sequence positions, from most important (0.0) to least important (100.0). This relative phylogenetic ranking reflects the variation entropy of each sequence position within and across the branches of an associated evolutionary tree, revealing evolutionary pressure points that correspond to functional and structural determinants, and the protein sites at which they often cluster (Mihalek et al., 2004). Many studies validated this approach for predicting binding and catalytic functional sites (Lichtarge et al., 1996; Onrust et al., 1997), guiding protein engineering (Peterson et al., 2015; Shenoy et al., 2006; Sowa et al., 2001, 2000) and predicting function (Amin et al., 2013). When viewed as gradient of the evolutionary landscape, ET rankings of residue importance can be combined with amino acid substitution log odds to estimate the likely impact, or Evolutionary Action (EA), of coding variations on protein function (Katsonis and Lichtarge, 2019, 2017, 2014). Here, this first ET and EA analysis of a full viral proteome identifies evolutionary important residues and functional sites in SARS-CoV-2.

RESULTS

Evolutionary Trace of SARS-CoV-2. In order to map functional sites and determinants in SARS-CoV-2 proteins we applied ET. With the multiple sequence alignments (**Figure S1A, Dataset S1**) and the corresponding phylogenetic trees (**Figure S2-S4**) in hand for 24 of the 26 SARS-CoV-2 proteins (see SI Methods and Materials), our protocol calculated the ET ranking of importance for 99.5% of SARS-CoV-2 amino acid residue positions (**Dataset S2**) generated from each of three protein databases (UniRef90, UniRef100, NCBI NR) and combined them into a single average. To independently assess the quality of these ranks, rather than rely on the variety and breadth of sequences in the alignments as indicative of information content, we used a statistical measure that quantifies the distribution of ET rankings in the 3D structure; the selection Cluster Weighting (SCW) z-score (Mihalek et al., 2004). This metric measures how well top-ranked ET residues cluster structurally relative to a randomized distribution of scores on the structure (see SI Materials and Methods). In previous studies, residues with better ET rankings (closer to 0) tended to cluster together at active sites, protein-protein interaction sites or other functional sites (Lichtarge et al., 1996; Mihalek et al., 2007, 2004; Wilkins et al., 2013, 2010). Here, such clustering of top-ranked residues was particularly prominent in several SARS-CoV-2 proteins and complexes including the NSP5 main protease, the NSP7/NSP8/NSP12 RNA-dependent RNA polymerase complex and the NSP10/NSP16 RNA cap methyltransferase complex and can be visualized as groups of warm colored residues in the protein structure (**Figure 1**). For almost all proteins, the SCW *z*-score is 2 standard deviations above the randomized background, confirming that the alignments are informative and that the resulting ET rankings are meaningful (**Figure S1, Dataset S3**). For the proteins that do not reach significant *z*-scores there is a clear correlation to a lack of sequences in the alignments (e.g. NSP1, E, ORF3, and ORF7a), or, the structure belongs to a small domain within a larger protein (e.g. the macrodomain within NSP3 and the HR2 domain within the S protein).

To probe these smaller domains within large proteins we further investigated the ADP-ribose-phosphatase (ADPRP) subdomain and macro and papain-like protease (PL^{pro}) domains of NSP3. NSP3 was an intriguing case because top-ranked ET residues cluster well in its PL^{pro} domain but not in its macrodomain or in the ADPRP subdomain (**Dataset S3**). In order to better resolve ET rankings for NSP3, we generated new alignments, phylogenetic trees, and ET residue rankings for the subsequences specific to each NSP3 domain structure (see SI Materials and Methods). In this focused analysis, the PL^{pro} domain now yielded ~50% more sequences leading to a corresponding increase in the clustering of top-ranked residues (**Figure S5**). For the macrodomain and ADPRP subdomain, thousands of additional sequences spanning the three domains of life and distantly related viruses were included in the new data set which resulted in ET rankings that rivaled the significance of clustering in the PL^{pro} domain. The stark differences we find in the phylogenetic trees of specific NSP3 domains confirm previous observations of alternate domain configurations in different coronavirus genera and even within clades of betacoronavirus (Lei et al., 2018). The improvement in SCW *z*-score corresponds to a cluster of highly ranked ET residues within the ligand binding site of the macro domain and ADPRP subdomain (**Figure S5D and E**) which was missing in the analysis of the full NSP3 reference sequence. Having better resolved ET rankings in the NSP3 domains, we returned to the main data set to see how well ET rankings captured functional sites in other proteins.

Phylogenetically conserved ligand binding sites. A catalog of SARS-CoV-2 ligand binding sites could serve as a timely resource for prioritizing therapeutic targets. Previous studies have shown that evolutionary sequence information correlates well-enough with enzyme active sites so as to serve as 3D-templates for functional signatures (Amin et al., 2013) and identify allosteric sites (Bhat et al., 2020; Rodriguez et al., 2010). Here we used NSP12, NSP15 and NSP16 as examples to show how the evolutionary sequence information captured by ET can successfully predict ligand binding sites for virus proteins. NSP12 is an RNA dependent polymerase, NSP15 mediates the cleavage of both single- and double-stranded RNA at uridine sites (Ulferts and Ziebuhr, 2011) and NSP16 is a m7GpppA-specific, S-adenosylmethionine (SAM)-dependent, 2'-O-MTase (Decroly et

al., 2011). As shown in **Figure 2A-C**, top ranked ET residues cluster around the native ligands of NSP12 (RNA) (Liu et al., 2020), NSP15 (GpU) (Youngchang Kim et al., 2020) and NSP16 (m7GpppA and SAM) (Rosas-Lemus et al., 2020), indicating an accurate prediction of ligand binding sites for these proteins. In order to quantify this result, the enrichment of highly ranked ET residues (ET ranking <= 30) around the ligand binding sites (within 5Å of the ligand) of NSP3, NSP12, NSP13, NSP14, NSP15 and NSP16 was ascertained by a Fisher's exact test. In each case, there is a statistically significant enrichment (**Figure S6 and Dataset S4**). Using this statistical test, we also find that in nine of the eleven ligand binding sites analyzed, our analysis better identifies known ligand binding sites than a previously reported analysis of the SARS-CoV-2 proteome using a percent identity metric (Figure S6 and Table S4) (R. Gupta et al., 2020). Moreover, several new functional sites are also predicted by ET (**Figure 2D and 2E**). On the spike protein (S), one such ET cluster partially overlaps the S2' protease cleavage site and fusion peptide that is critical for membrane fusion and infectivity of the SARS virus (Madu et al., 2009). On the N-terminal domain of the nucleoprotein (N), a cluster of highly ranked ET residues are either adjacent to (Dinesh et al., 2020) or overlap (Lin et al., 2014; Saikatendu et al., 2007) the putative RNA binding site and may contribute to formation of the N protein-RNA helical filaments that package the RNA genome (Chen et al., 2007). These results indicate ET can provide alternative drug target sites with no currently available ligand-bound structures.

In addition to being important to protein function, ideal drug target sites should also be rarely mutated in the current outbreak so as to avoid the potential emergence of drug resistance. Thus, we focused on positions that do not have any mutations observed in the 139,607 high quality, full length SARS-CoV-2 sequences that were available as of December 8th, 2020. As more genomes and mutations within them are sequenced it may be necessary to lower the variant count stringency. In order to translate proteome-wide ET ranks and mutational profiles into useful information for drug development, we defined clusters of mutation-free, surface-exposed residues that are highly ranked by ET and fall within 5Å of each other (**Figure 3**, **Dataset S5**) as variant adjusted 3D sites. The resulting catalog of potential drug targets includes 103 sites at ~4 sites per

structure with the largest structure (full-length model of Spike, 6vsb_1_1_1) having the highest number of sites. For NSP12, NSP15 and NSP16, the variant adjusted 3D sites overlap the known ligand binding sites.

In order to evaluate whether these variant adjusted 3D sites may correspond to druggable target sites, we examined their overlap with sites observed in five SARS-CoV-2 protein-drug complex crystal structures. It is important to note that all 5 drugs showed an inhibitory effect in either cellular or biochemical assays. Remdesivir has been shown to speed up the recovery of COVID-19 patients in clinical trials (Beigel et al., 2020), while the α-ketoamide inhibitor 13b can suppress SARS-CoV-2 replication in cell lines (L. Zhang et al., 2020). Vir251 and tipiracil were also shown to effectively inhibit the enzymatic activities of their targets (Youngchang Kim et al., 2020; Rut et al., 2020). The remaining drug, sinefungin, is a pan-MTnase (NSP16) inhibitor that inhibits the growth of yeast cells ectopically expressing NSP16 from SARS-CoV (Decroly et al., 2011). The variant adjusted 3D sites were mapped onto the five SARS-CoV-2 protein-drug complexes (Youngchang Kim et al., 2020; Minasov et al., 2020a; Rut et al., 2020; Yin et al., 2020; L. Zhang et al., 2020) and, as shown in **Figure 3**, all five drugs reside in protein surface pockets that are within or are very close to at least one residue of a predicted variant adjusted 3D site. The variant adjusted 3D site for NSP5 is not well recovered mostly due to a single SARS-CoV-2 sequencing entry (strain MT745875) wherein several residues in the protease active site are mutated (G143S, S144E and C145I), including the catalytic cystine residue. S144E and C145I are both caused by two nucleotide substitutions in the codon, and only observed in this strain (sampled on 06/24/20). It is unclear whether this is a sequencing artifact or represents a genuine active site plasticity that compromises NSP5's active site as a stable drug target. It does however illustrate the importance of accurately detecting emerging sequence variations when choosing drug targets. A clearer example of this is the tipiracil bound NSP15 active site which, although evolutionarily important to the overall coronavirus family, is not predicted to be a good drug target due to the presence of multiple variants observed in the current outbreak. Overall, these results show that predicted variant adjusted 3D sites can recover experimentally tested drug binding pockets and suggest new sites that can be

targeted in computational docking approaches. In addition, because these sites are conserved across multiple coronavirus genera, these predicted variant adjusted 3D sites are anticipated to be relevant for identifying inhibitors of SARS-CoV-2 as well as more distantly related coronaviruses.

Conserved linear sites. Variant adjusted 3D sites may prove valuable in guiding drug design, but these approaches are dependent upon having high resolution crystal structures and some structures are either not yet available (e.g. NSP2, NSP6, M, and several accessory proteins), do not cover a majority of the protein (NSP3 and NSP4) or are too low in resolution for accurate docking studies (NSP12, NSP14, ectodomain of S, N, ORF3a and ORF7a). However, ET operates over protein sequences and can therefore identify phylogenetically important linear sequence fragments even in the absence of a 3D structure (Lichtarge et al., 2002). As in our approach to discover variant adjusted 3D sites, we combined ET residue ranking information with sequencing data from SARS-CoV-2 isolates to arrive at linear peptides along the proteome that are evolutionarily important and also show no variation in the current outbreak (**Figure S7, Dataset S6**). In order to assess the value of these variant adjusted linear sites, we asked whether they could recapitulate variant adjusted 3D sites. Variant adjusted linear sites for NSP12 were mapped onto an available NSP12 structure and, as illustrated in **Figure 4A**, the majority of the 3D and linear sites overlap with each other. Variant adjusted linear sites and 3D sites overlap well for other SARS-CoV-2 proteins, which was quantified by Jaccard Similarity and Fisher's exact test (**Dataset S7**). These data suggest that variant adjusted linear sites contain functionally relevant information since they recapitulate variant adjusted 3D sites for proteins or domains without requiring 3D structural data. In the absence of a protein structure, these linear sites could be useful in designing inhibitory peptides (Gu et al., 2005; Wu et al., 2020).

These peptides are also connected to a second main approach towards resolving the pandemic, vaccine and mAb development. Although vaccines for COVID-19 are now available, several new variants that arose in the UK (B.1.1.7) (Rambaut et al., 2020) and South Africa (B.1.351, also known as 501Y.V2) (Tegally et al., 2020) have multiple substitutions in the Spike protein's receptor binding

domain. Both are resistant to several classes of mAbs (Ho et al., 2021; Wibmer et al., 2021). While B.1.1.7 is ~ 2 fold more resistant to convalescent plasma, B.1.351 is more concerning as it can be ~11-33 fold more resistant to the convalescent plasma obtained from ~80% of patients (Ho et al., 2021; Wibmer et al., 2021). Ideally, effective protection against future outbreaks from related coronaviruses would include a broadly neutralizing effect wherein the immune system recognizes epitopes shared among coronavirus species. The prospect of raising a broad antibody response is bolstered by a study that naïve patients, never exposed to SARS-CoV-2, were found to possess a subset of T-cells that can cross-react to homologous epitopes shared by common cold coronaviruses and SARS-CoV-2 (Mateus et al., 2020). In this context, we note that ET rankings reflect the degree of homology over the phylogenetic tree, so we reasoned that summing ET scores over the length of an identified T-cell epitope may be able to estimate its potential for cross-reactivity.

As a first step, we summed the ET ranks for each of the 40 SARS-CoV-2 epitopes that had been shown to react with patient-derived T-cells so that they could be ranked by predicted cross-reactivity to 161 common cold coronavirus epitopes assayed by Mateus et al. Although summing ET ranks could identify SARS-CoV-2 epitopes that are more likely to be cross-reactive (**Figure S8**), it did not account for the specific amino acid differences in the potentially cross-reactive homolog. In other words, ET ranks can predict whether or not a SARS-CoV-2 epitope will be cross-reactive in general, but not which epitope homologs will cross react.

In order to improve resolution of our predictions to specific epitope homologs, we next combined EA, a predictor of mutational impact, with the summed ET rankings. EA calculates the predicted impact of amino acid variations on protein function aiding in the interpretation of coding variants (Katsonis and Lichtarge, 2019, 2017, 2014). Summing the predicted impact of amino acid changes between a SARS-CoV-2 epitope and a homologous epitope in another virus (sumEA) while adjusting for the SARS-CoV-2 epitope's overall evolutionary importance (sum(100-ET ranking)) produced a metric that was able to separate cross-reactive epitopes from those that did not cross

react (**Figure 4B and S8, Dataset S8**). This metric, sumEA/sum(100-ET ranking), was then applied to 21 untested SARS-CoV-2 T-cell epitopes and their common cold homologs (Mateus et al., 2020). From a total of 92 homologs we identified 23 with potential to cross react to one of five SARS-CoV-2 epitopes (**Figure 4C, Dataset S9**). These 5 SARS-CoV-2 epitopes along with the 9 others experimentally shown to possess cross-reactivity could be used in a multi-epitope vaccination strategy that provides a broad neutralizing response to currently circulating coronaviruses, SARS-CoV-2 and, possibly, future outbreaks. Subsequent to this analysis, further confirmation came when it was found that residues 815-825 of the Spike protein compose the most frequently recognized epitope among naïve and COVID-19 patients (Shrock et al., 2020). These 11 residues are specifically highlighted by ET as being particularly evolutionarily conserved amino acids and are thereby responsible for our metric's prediction of cross reactivity in the 15 amino acid long epitopes used in the study by Mateus et al. This result and the generality of our approach suggest highly cross-reactive epitopes could be quickly identified in other families of pathogens.

Dissemination. To disseminate these results, a public website (http://cov.lichtargelab.org) makes these data and analyses fully accessible. The data include, for example, multiple sequence alignments, pre-calculated ET ranks, and predicted sites (both linear and structural) for all SARS-CoV-2 proteins. In addition, an interactive structure viewer enables users to explore ET-colored structures and predicted variant adjusted ET sites associated with those structures (**Dataset S5-6**). The website will be updated as new SARS-CoV-2 isolates and protein structures become available.

DISCUSSION

Rapid progress has been made in response to the SARS-CoV-2 pandemic; from sequencing, to structural determination, and drug and vaccine development (Jeong et al., 2020; Kneller et al., 2020; Li et al., 2020; Sanders et al., 2020; Wu et al., 2020). Here, we make use of coronavirus phylogenetics, and sequence and structure information to provide a functional map of sites that are

not only stable across coronavirus families but are also stable to mutations in the current pandemic. These sites are favored strategic targets for pan coronavirus/betacoronavirus therapeutics that are less likely to be subjected to the rapid emergence of resistance from SARS-CoV-2 variants. In addition to focusing therapeutic studies, the data presented here can guide directed mutagenesis studies aimed at identifying the mechanism of action for successful therapies, not only in the context of the current outbreak but across future coronavirus outbreaks.

There are limitations to this study. The quality of our results depends on the number and range of homologous sequences available and a few of the SARS-CoV-2 proteins such as NSP1 and the accessory proteins do not reach significant z-scores or have many diverse sequences in their final alignments. The inability to recover more sequence information could be due to a higher evolutionary rate in these proteins that limits our ability to recognize distantly related homologs with very little sequence identity. More likely, these peripheral genes have been more recently recruited through the frequent recombination events that occur in the coronavirus family (Su et al., 2016). The equivalent of gene recruitment has occurred at the domain level in the NSP3 protein with its variable number of domains (10 to 16), some of which are unique to the betacoronavirus clade b containing SARS-CoV and SARS-CoV-2. Therefore, it is unsurprising that the initial alignments and corresponding ET rankings for full-length NSP3 are heavily influenced by the less divergent PL^{pro} domain that is present across coronavirus clades and families. Domain-specific analysis of NSP3 greatly improved both the number of sequences returned, phylogenetic coverage, and the resolution of ET results. This suggests that future work should include domain specific analyses for multidomain proteins. Such analyses are likely to provide ET rankings that identify important functional sites for individual domains, while full-length analyses can provide insight into how particular domains were recruited in specific branches of the phylogenetic tree.

Several other groups have focused on experimentally screening clinical-stage or FDA-approved small molecules with the hope of identifying and repurposing drugs for SARS-CoV-2 treatment. However, drug efficacy of top hits might be cell line specific (Hoffmann et al., 2020) and the

mechanisms of drug action may be unclear or acting through modulation of the tissue culture cell. In silico docking studies (Deshpande et al., 2020; S. Gupta et al., 2020) take a more targeted approach towards specific SARS-CoV-2 sites and benefit from knowledge of ligand binding sites. Although structural characterization of SARS-CoV-2 proteins is unprecedented, the structural information available is far from comprehensive. In order to bridge these knowledge gaps, we identified 3D clusters of surface residues that have low ET rankings and a lack of mutations in the current outbreak as potential drug target sites. Many of these variant adjusted 3D sites correspond to ligand bound active sites, but others map to evolutionarily important sites that have yet to be fully characterized. These variant adjusted sites are putative drug targets which can guide docking studies to sites not immediately apparent from currently available structural information.

The depth of the phylogenetic tree for the sequences analyzed with ET can set expectations for how broadly a drug may inhibit homologs in different species. For instance, the active site of NSP12 is conserved throughout a deep phylogenetic tree of RNA viruses and an inhibitor targeting it, remdesivir, is effective against SARS-CoV-2, SARS-CoV, MERS and the distantly related Ebola RNA virus (de Wit et al., 2020; Eastman et al., 2020). Likewise, most of the other NSPs and the structural M (membrane) and N (nucleocapsid) proteins have deep RNA virus phylogenies and targeting them may provide broadly effective inhibitors. The most obvious candidates are the variant adjusted 3D sites that overlap with the ligand binding sites of NSP12, NSP13, NSP14 and 16. A less apparent drug target revealed by our analysis includes a putative RNA binding site on the N terminal domain of the N protein (Kang et al., 2020). The equivalent site in HCoV-OC43's N protein is targeted by compound PJ34 where it inhibits RNA binding activity and viral replication (Lin et al., 2014; Peng et al., 2020). In contrast, another inhibitor (5-benzyloxygramine), that induces aggregation of the MERS-CoV N protein and shows potent antiviral activity against that virus (Lin et al., 2020), binds two hydrophobic pockets. However, these pockets have undergone variation in the current SARS-CoV-2 outbreak, suggesting

compounds targeting these areas are more likely to become susceptible to resistant SARS-CoV-2 variants.

In contrast to the aforementioned deep phylogenies predominantly composed of RNA virus sequences, the ADP ribose phosphatase sub-domain of NSP3 has a phylogenetic tree with few coronavirus sequences among a multitude of sequences that span three domains of life. Drugs targeting this domain may inhibit coronavirus infectivity but could also inhibit host ADP ribose phosphatases. ADP ribose phosphatase inhibitors have already been developed for cancer treatment and their application towards SARS-CoV-2 treatment is warranted (Kassab et al., 2020) but care should be taken to ensure unwanted side effects do not overshadow any benefits as a viral inhibitor.

Along with small molecule viral inhibitors, the development of immunological therapeutics to address COVID-19 can also be guided by the use of evolutionary information. We performed evolutionary analysis on SARS-CoV-2 T-cell epitopes capable of cross reacting with homologous peptides in other human coronaviruses (Le Bert et al., 2020; Mateus et al., 2020). This led to a new metric, sumEA/sum(100-ET ranking), that can better predict which epitopes will cross-react. In general, knowledge of cross-reactive epitopes could inform multi-epitope vaccine development efforts to direct the immune system towards a broadly neutralizing response.

The S protein evolves to bind different receptors (Fehr and Perlman, 2015), suggesting the high variation rate of the binding site is due to both this adaptive function and the avoidance of the host's adaptive immune system. Though several vaccines and mAbs are now in use, new strains are arising rapidly (Faria et al., 2021; Fiorentini et al., 2021; Rambaut et al., 2020; Tegally et al., 2020) and show signs of evading existing treatments (Wibmer et al., 2021; Wu et al., 2021). The two most worrisome mutations in those strains occur at N501 and E484 in the receptor binding domain of the S protein. These residues have ET ranks of 96 and 95.8 respectively, indicative of particularly rapid phylogenetic change across the coronavirus family and represent poor targets for therapeutics meant to remain effective against emerging SARS-CoV-2 variants. Despite the

variability of the S protein, we identified a relatively conserved site (**Figure 2D**) that corresponds to a fusion peptide adjacent to the S2' cleavage site (Madu et al., 2009) that is also the most cross-reactive epitope among naïve and COVID-19 patients (Shrock et al., 2020) (**Figure 4C**). This site is particularly appealing when considering that the 5H10 human mAb targeting the equivalent region in SARS-CoV was very effective in preventing disease a *Rhesus macaque* infection model (Miyoshi-Akiyama et al., 2011). The appearance of new strains makes it very likely that additional vaccine and mAb development will be necessary. We believe that targeting the SARS-CoV-2 variant-adjusted linear ET site corresponding to the S2' cleavage site and fusion peptide region and other sites highlighted by our study, may provide protection that is less susceptible to the emergence of resistant variants.

CONCLUSION

This study was motivated by the current pandemic and uses evolutionary sequence information to guide the development of therapeutics for COVID-19. Although we are presently in the grip of COVID-19, this pandemic was preceded by the SARS and MERS outbreaks and it should be anticipated that related coronaviruses will cause future outbreaks. And while this study is focused upon SARS-CoV-2, it draws upon pieces of sequence information taken from the whole of the coronavirus family and thereby the findings are extendable to other coronavirus species, including those that have not yet been encountered. Indeed, the tools we present could be applied to any family of pathogen. Putting a pandemic virus into the evolutionary context of related viruses can expose a path to managing recovery and may offer therapeutics that cover future outbreaks.

MATERIALS AND METHODS

A brief description of the methods can be found here; for a more in-depth description of specific methods please see the Supplementary text.

Evolutionary Trace:

In order to map functional determinants in SARS-CoV-2 proteins we applied the Evolutionary Trace (ET) approach (Lichtarge et al., 1996; Mihalek et al., 2004). This method ranks each amino acid position from most to least important during evolution by tracking how they vary along the coronavirus phylogenetic tree. These rankings vary based on the precise choice of multiple sequence alignment (MSA). In order to produce robust ET rankings three separate alignments were generated for each protein in the SARS-CoV-2 Wuhan-Hu-1 reference genome (NC_045512.2) (Wu et al., 2020), by querying three protein databases (UniRef90, UniRef100, and NCBI NR) for sequences with identity between 25% and 98%. This procedure filtered out sequences that were either overly distant or redundant. Only two proteins had too few matches for ET, NSP11 and ORF10, both of which have unknown function (Gadhave et al., 2021; Pancer et al., 2020) and have very short reference sequences (13 and 38 amino acids, respectively, **FigureS1, Dataset S1**). The ET scores for all other proteins for each alignment and for the average scores across alignments were evaluated with the previously presented Selection Cluster Weighting (SCW) *z*-score (Mihalek et al., 2007, 2004; Wilkins et al., 2013, 2010). The *z*-scores for each structure were then ranked 1-4 in order to determine if ET scores from one database or the average of the three consistently outperforms the others. ET scores from each of the three databases performed similarly well but the average ET of the three provided better *z*-scores in most cases (**Figure S1C**). ET rankings were further investigated by comparing the highest scoring regions with known functional sites.

Prediction of Variant Adjusted ET Sites:

Variant adjusted ET sites were predicted based on both the linear sequence as well as structural constraints. Residues were nominated as members of potential therapeutic sites based on their

ET rankings, lack of variants as found in SARS-CoV-2 sequences retrieved from GISAID (Shu and McCauley, 2017), Genbank (Benson et al., 2018), and the China National Center for Bioinformation (CNCB) (Zhao et al., 2020), as well as surface accessibility, and structural proximity. Structurally identified therapeutic sites were compared to drug binding sites for agents known to bind to SARS-CoV-2 proteins. To generalize this approach to proteins without structure, linear sites were predicted based on ET rankings, current mutational profile and linear connectivity. Structural and linear predicted sites were compared to one another using Jaccard Similarity and Fisher's Exact test, to determine the usefulness of this method in the absence of a protein structure. Several ET metrics were also interrogated to determine their ability to highlight potential cross-reactive immunogenic epitopes (Mateus et al., 2020). The best metric, sumEA/sum(100-ET ranking), was used to predict cross-reactive T-cell epitopes which are good potential therapeutic sites.

Acknowledgments

This research is based upon work supported [in part] by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) under BAA-17-01, contract #2019-19071900001. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. The authors also gratefully acknowledge support from the National Science Foundation (DBI-2032904), the Oskar Fischer Foundation, and the National Institutes of Health (GM079656, GM066099, and AG061105).

Conflict of interest statement

The authors of this text have no conflicts of interest to report.

Figure 1

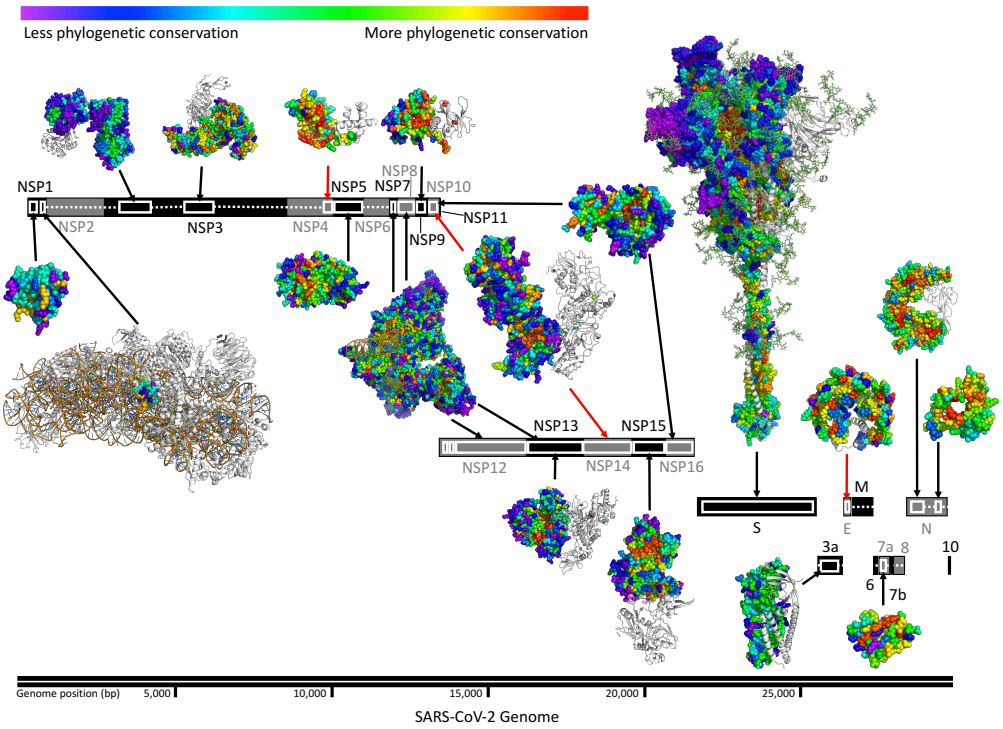

Figure 1. Structural and sequence information permits identification of evolutionarily important sites in SARS-CoV-2. Linear representation of SARS-CoV-2 proteome with structurally determined regions highlighted (white boxes) and corresponding structures' residues colored by Evolutionary Trace rank. Black arrows connect SARS-CoV-2 structures to their corresponding gene, red arrows indicate that only structures of homologous proteins are available. Host ribosomal proteins in the NSP1 complex structure are shown in white. For multimeric structures, one monomer is also shown in white. Structures shown include: 7k7p (Clark et al., 2020) (NSP1), 6zlw (Thoms et al., 2020) (NSP1 C-term), 6woj (Alhammad et al., 2020) (NSP3), 6w9c (Osipiuk et al., 2020) (NSP3), ExPasy NSP4 model 01 (Bienert et al., 2017; "Non-structural protein 4 (nsp4) | P0DTD1 PRO_0000449622 | Models," n.d.; Studer et al., 2020; Waterhouse et al., 2018) (NSP4), 6yb7 (Owen et al., 2020) (NSP5), 6wxd (Littler et al., 2020) (NSP9), 6xez (Chen et al., 2020) (NSP7, 8 , 12, and 13), 6zsl (Newman et al., 2020) (NSP13), 5c8s (Ma et al., 2015) (NSP10 and 14), 6wlc (Y. Kim et al., 2020) (NSP15), 6w4h (Minasov et al., 2020b) (NSP10 and NSP16), 6vsb_1_1_1 (Woo et al., 2020) (S), 6xdc (Kern et al., 2020) (ORF3a), 5x29 (Surya et al., 2018) (E), 6w37 (Nelson et al., 2020) (ORF7a), 6vyo (Chang et al., 2020) (N), and 6zco (Zinzula et al., 2020) (N).

Figure 2

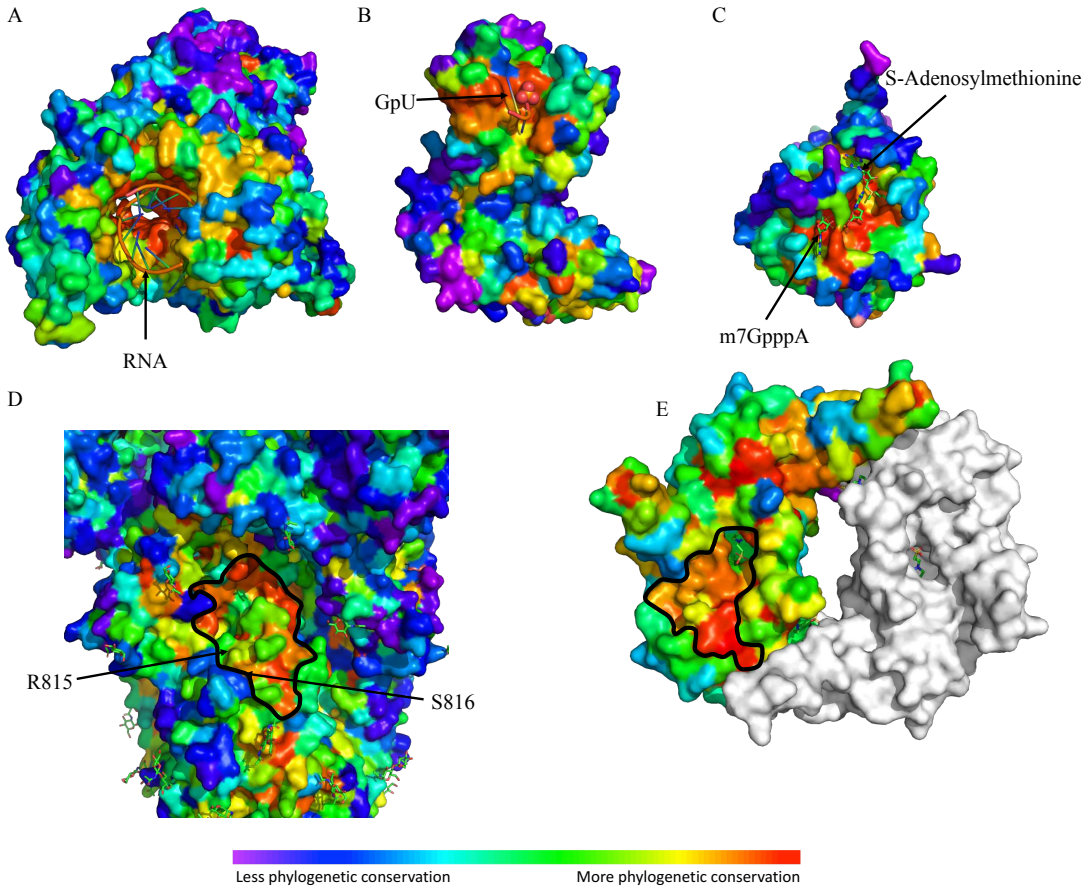

Figure 2. Top ET ranking residues overlap with known functional sites. ET recovers (A) the RNA binding site of NSP12 (RNA dependent RNA polymerase, pdb:6xez), (B) the active site of NSP15 (uridine-specific endoribonuclease, pdb:6x1b), (C) the substrates binding sites of NSP16 (RNA-cap methyltransferase, pdb:6wvn). ET also recovers a key functional (D) S2' protease cleavage site of S (key residues: R815 and S816, pdb:6vsb), and predicts (E) a site associated with the putative RNA binding site of N (pdb:6vyo). The S cleavage site between R815 and S816 is labeled and the putative sites of panels D and E are highlighted with a black outline. The site in panel D is a high priority target as it was found that residues 815-825, which overlap this site, comprise the most frequently recognized epitope among naïve and COVID-19 patients (Shrock et al., 2020).

Figure 3

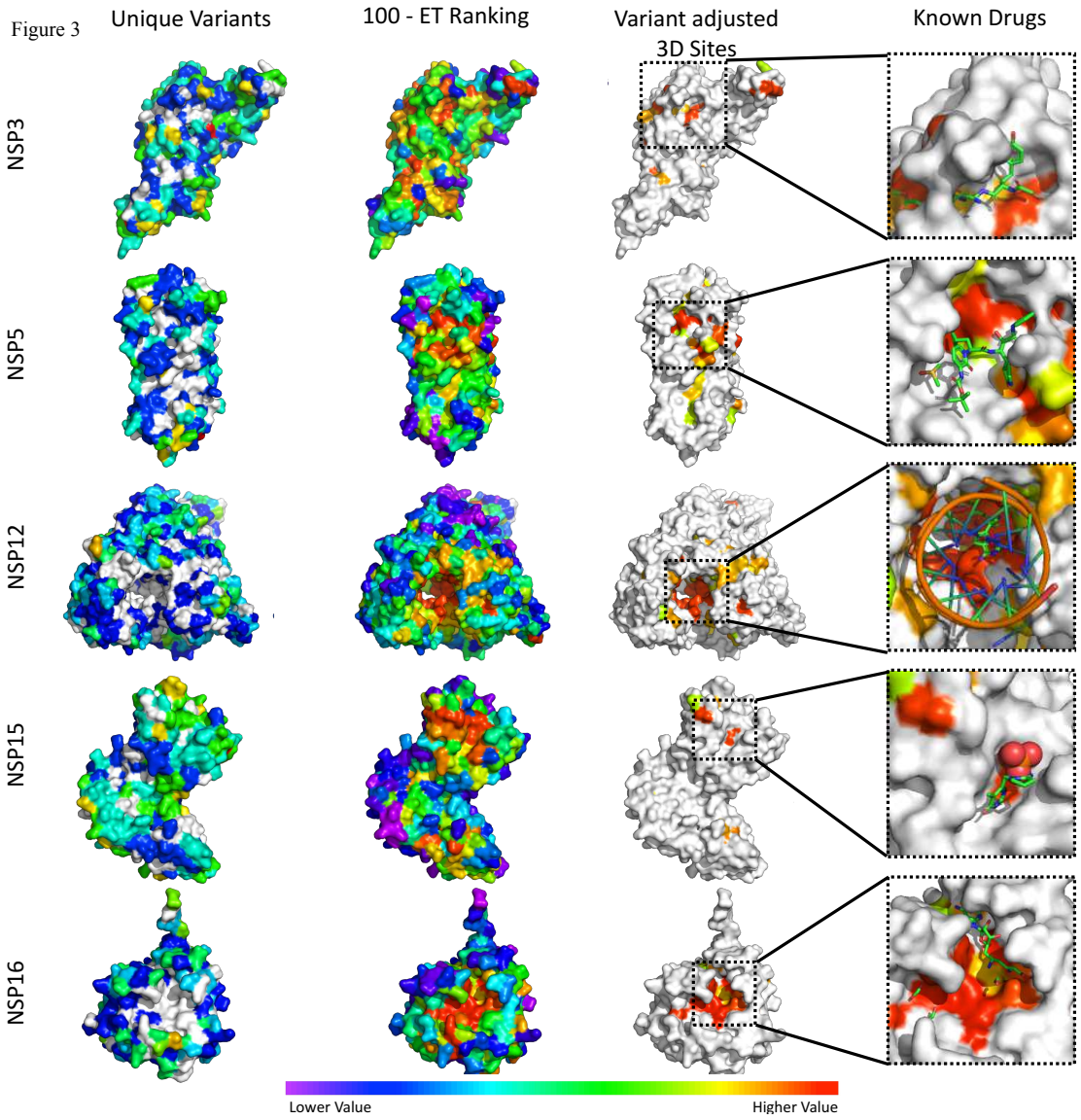

Figure 3. Identification of variant adjusted 3D sites (5Å) and their colocalization with known drug binding sites. Variant adjusted 3D sites for NSP3 (6w9c(Osipiuk et al., 2020)), NSP5 (6yb7(Owen et al., 2020)), NSP12 (7bv1(Yin et al., 2020)), NSP15 (6wlc(Youngchang Kim et al., 2020)), and NSP16 (6w4h(Minasov et al., 2020b)) were identified as clusters of surface residues with low ET ranks and a lack of mutations in the current outbreak. In the known drugs panels, variant adjusted 3D sites were identified using apo form structures, then mapped to the co-structures of NSP3 with peptide inhibitor vir251 (PDB:6wx4(Rut et al., 2020)), NSP5 with potential drug 13b (PDB:6y2f(L. Zhang et al., 2020)), NSP12 with drug remdesivir (7bv2(Yin et al., 2020)), NSP15 in complex with potential drug tipiracil (PDB:6wxc(Youngchang Kim et al., 2020)), and NSP16 with sinefungin (PDB:6wkq(Minasov et al., 2020a)). For structures in the "Unique Variants" column "Lower Values" in the color scale correspond to fewer variants, while "Higher Values" correspond to more variants and white residues have no reported variants in the analyzed SARS-CoV-2 strains. For the "100 – ET Ranking", "Variant Adjusted 3D Sites", and "Known Drugs" columns, "Lower Values" correspond to less phylogenetic conservation while "Higher Values" correspond to more phylogenetic conservation.

Figure 4

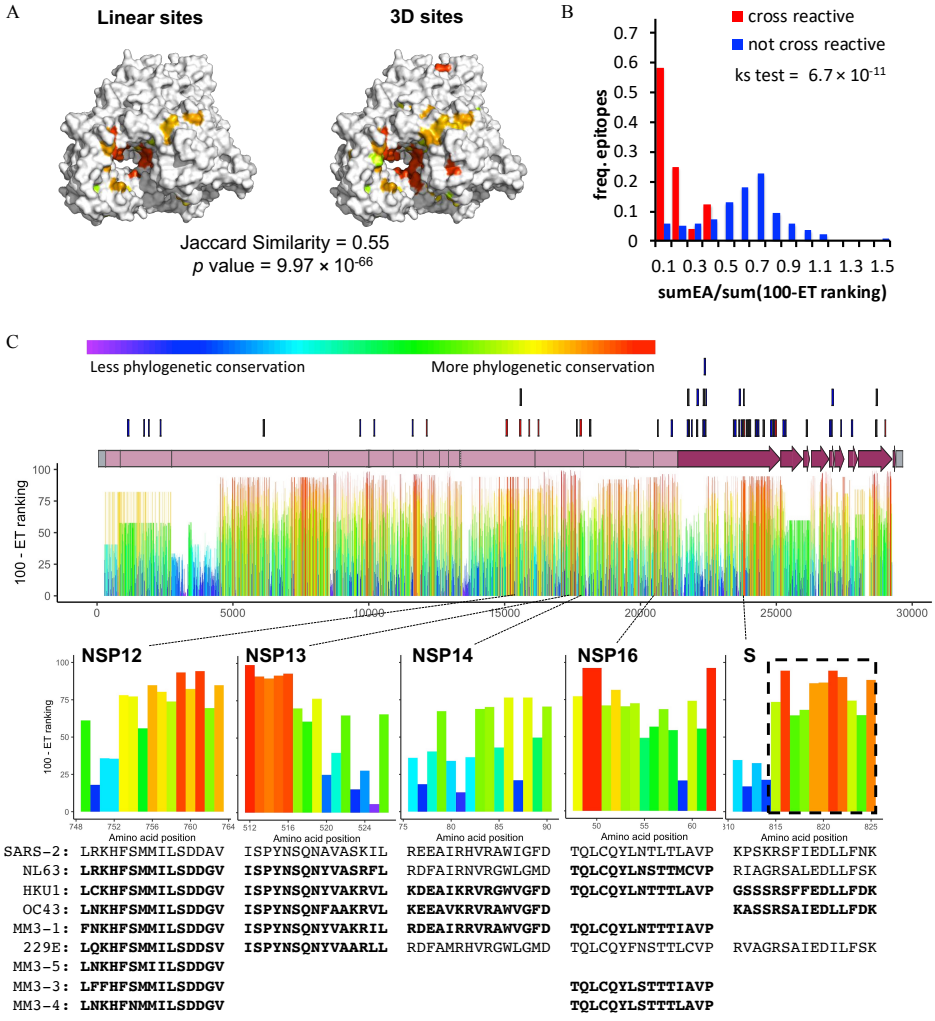

Figure 4. Identification of variant adjusted linear sites with Evolutionary Trace. A) Mapping of variant adjusted linear and structural sites on the surface of NSP12 (7bv1) with Jaccard Similarity value and Fisher's Exact test *p* value indicated. B) Relative frequency distributions of sumEA/sum(100-ET ranking) for T-cell epitopes shown to either be cross reactive (red) or not (blue). The sumEA/sum(100-ET ranking) metric predicts the functional impact of variants (EA) relative to the overall Evolutionary Trace rankings in the epitope. A Kolmogorov-Smirnov test (ks test) shows a significant difference in the distributions. C) T-cell epitopes reported in Mateus, et al. (Mateus et al., 2020) are shown above the SARS-CoV-2 genome (lines) and ET rankings within each protein are shown below. Shown are five SARS-CoV-2 epitopes (NSP12, NSP13, NSP14, NSP16 or S) that are predicted to cross react with the indicated common human coronavirus epitopes (bold text). Closely related coronavirus epitopes that did not meet our stringent threshold are also shown (normal text). The dashed box highlights the 11 amino acid stretch subsequently shown to be the most cross-reactive Spike protein epitope among naïve and COVID-19 patients (Shrock et al., 2020).

References

Ahmed SF, Quadeer AA, McKay MR. 2020. Preliminary identification of potential vaccine targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* **12**:254. doi:10.3390/v12030254

Alhammad YMO, Kashipathy MM, Roy A, Gagné J-P, Nonfoux L, McDonald P, Gao P, Battaile KP, Johnson DK, Poirier GG, Lovell S, Fehr AR. 2020. The SARS-CoV-2 conserved macrodomain is a highly efficient ADP-ribosylhydrolase. *bioRxiv* 2020.05.11.089375. doi:10.1101/2020.05.11.089375

Amin SR, Erdin S, Ward RM, Lua RC, Lichtarge O. 2013. Prediction and experimental validation of enzyme substrate specificity in protein structures. *Proc Natl Acad Sci U S A* **110**:E4195-202. doi:10.1073/pnas.1305162110

Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, Diemert D, Spector SA, Rouphael N, Creech CB, McGettigan J, Khetan S, Segall N, Solis J, Brosz A, Fierro C, Schwartz H, Neuzil K, Corey L, Gilbert P, Janes H, Follmann D, Marovich M, Mascola J, Polakowski L, Ledgerwood J, Graham BS, Bennett H, Pajon R, Knightly C, Leav B, Deng W, Zhou H, Han S, Ivarsson M, Miller J, Zaks T, COVE Study Group. 2021. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* **384**:403–416. doi:10.1056/NEJMoa2035389

Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, Hohmann E, Chu HY, Luetkemeyer A, Kline S, Lopez de Castilla D, Finberg RW, Dierberg K, Tapson V, Hsieh L, Patterson TF, Paredes R, Sweeney DA, Short WR, Touloumi G, Lye DC, Ohmagari N, Oh M-D, Ruiz-Palacios GM, Benfield T, Fätkenheuer G, Kortepeter MG, Atmar RL, Creech CB, Lundgren J, Babiker AG, Pett S, Neaton JD, Burgess TH, Bonnett T, Green M, Makowski M, Osinusi A, Nayak S, Lane HC, ACTT-1 Study Group Members. 2020. Remdesivir for the Treatment of Covid-19 - Preliminary Report. *N Engl J Med*. doi:10.1056/NEJMoa2007764

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW. 2018. GenBank. *Nucleic Acids Res* **46**:D41–D47. doi:10.1093/nar/gkx1094

Bhat AS, Dustin Schaeffer R, Kinch L, Medvedev KE, Grishin N V. 2020. Recent advances suggest increased influence of selective pressure in allostery. *Curr Opin Struct Biol*. doi:10.1016/j.sbi.2020.02.004

Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC, Lee SS, Chakraborty C. 2020. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *J Med Virol* **92**:618–631. doi:10.1002/jmv.25736

Bienert S, Waterhouse A, De Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. 2017. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res* **45**:D313–D319. doi:10.1093/nar/gkw1132

Chan JF-W, Kok K-H, Zhu Z, Chu H, To KK-W, Yuan S, Yuen K-Y. 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* **9**:221–236. doi:10.1080/22221751.2020.1719902

Chang C, Michalska K, Jedrzejczak R, Maltseva N, Endres M, Godzik A, Kim Y, Joachimiak A, Center for Structural Genomics of Infectious Diseases (CSGID). 2020. RCSB PDB - 6VYO: Crystal structure of RNA binding domain of nucleocapsid phosphoprotein from SARS

coronavirus 2. doi:10.2210/pdb6VYO/pdb

Chen CY, Chang C ke, Chang YW, Sue SC, Bai HI, Riang L, Hsiao CD, Huang T huang. 2007. Structure of the SARS Coronavirus Nucleocapsid Protein RNA-binding Dimerization Domain Suggests a Mechanism for Helical Packaging of Viral RNA. *J Mol Biol* **368**:1075–1086. doi:10.1016/j.jmb.2007.02.069

Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi K, Eng ET, Vatandaslar H, Chait BT, Kapoor TM, Darst SA, Campbell EA. 2020. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*. doi:10.1016/j.cell.2020.07.033

Clark LK, Green TJ, Petit CM. 2020. Structure of Nonstructural Protein 1 from SARS-CoV-2. *J Virol* **95**. doi:10.1128/jvi.02019-20

de Wit E, Feldmann F, Cronin J, Jordan R, Okumura A, Thomas T, Scott D, Cihlar T, Feldmann H. 2020. Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. *Proc Natl Acad Sci U S A* **117**:6771–6776. doi:10.1073/pnas.1922083117

Decroly E, Debarnot C, Ferron F, Bouvet M, Coutard B, Imbert I, Gluais L, Papageorgiou N, Sharff A, Bricogne G, Ortiz-Lombardia M, Lescar J, Canard B. 2011. Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2′-O-Methyltransferase nsp10/nsp16 Complex. *PLoS Pathog* **7**:e1002059. doi:10.1371/journal.ppat.1002059

Deshpande RR, Tiwari AP, Nyayanit N, Modak M. 2020. In silico molecular docking analysis for repurposing therapeutics against multiple proteins from SARS-CoV-2. *Eur J Pharmacol* **886**:173430. doi:10.1016/j.ejphar.2020.173430

Dhama K, Sharun K, Tiwari R, Dadar M, Malik YS, Singh KP, Chaicumpa W. 2020. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. *Hum Vaccin Immunother* **16**:1232–1238. doi:10.1080/21645515.2020.1735227

Dinesh DC, Chalupska D, Silhan J, Veverka V, Boura E. 2020. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *bioRxiv* 2020.04.02.022194. doi:10.1101/2020.04.02.022194

Dong E, Du H, Gardner L. 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. doi:10.1016/S1473-3099(20)30120-1

Eastman RT, Roth JS, Brimacombe KR, Simeonov A, Shen M, Patnaik S, Hall MD. 2020. Remdesivir: A Review of Its Discovery and Development Leading to Emergency Use Authorization for Treatment of COVID-19. *ACS Cent Sci* **6**:672–683. doi:10.1021/acscentsci.0c00489

Faria NR, Claro IM, Candido D, Franco LAM, Andrade PS, Coletti TM, Silva CAM, Sales FC, Manuli ER, Aguiar RS, Gaburo N, Camilo C da C, Fraiji NA, Crispim MAE, Carvalho M do PSS, Rambaut A, Loman N, Pybus OG, Sabino EC. 2021. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586

Fehr AR, Perlman S. 2015. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* **1282**:1–23. doi:10.1007/978-1-4939-2438-7_1

Fiorentini S, Messali S, Zani A, Caccuri F, Giovanetti M, Ciccozzi M, Caruso A. 2021. First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020. *Lancet Infect Dis*. doi:10.1016/S1473-3099(21)00007-4

Gadhave K, Kumar P, Kumar A, Bhardwaj T, Garg N, Giri R. 2021. Conformational Dynamics of NSP11 Peptide of SARS-CoV-2 Under Membrane Mimetics and Different Solvent Conditions. *bioRxiv* 2020.10.07.330068. doi:10.1101/2020.10.07.330068

Goodsell DS, Sanner MF, Olson AJ, Forli S. 2020. The AutoDock suite at 30. *Protein Sci* pro.3934. doi:10.1002/pro.3934

Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. 2020a. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **27**:671-680.e2. doi:10.1016/j.chom.2020.03.002

Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Jennifer M, Moderbacher CR, Rawlings SA, Sutherland A, Premkumar L, Jadi RS, Marrama D, Silva AM De, Frazier A, Carlin AF, Greenbaum JA, Peters B, Krammer F, Smith DM, Crotty S, Sette A, Dan JM, Moderbacher CR, Rawlings SA, Sutherland A, Premkumar L, Jadi RS, Marrama D, de Silva AM, Frazier A, Carlin AF, Greenbaum JA, Peters B, Krammer F, Smith DM, Crotty S, Sette A. 2020b. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* **181**:1489-1501.e15. doi:10.1016/j.cell.2020.05.015

Gu P, Morgan DH, Sattar M, Xu X, Wagner R, Raviscioni M, Lichtarge O, Cooney AJ. 2005. Evolutionary trace-based peptides identify a novel asymmetric interaction that mediates oligomerization in nuclear receptors. *J Biol Chem* **280**:31818–29. doi:10.1074/jbc.M501924200

Gupta M, Sharma R, Kumar A. 2018. Docking techniques in pharmacology: How much promising? *Comput Biol Chem* **76**:210–217. doi:10.1016/j.compbiolchem.2018.06.005

Gupta R, Charron J, Stenger CL, Painter J, Steward H, Cook TW, Faber W, Frisch A, Lind E, Bauss J, Li X, Sirpilla O, Soehnlen X, Underwood A, Hinds D, Morris M, Lamb N, Carcillo JA, Bupp C, Uhal BD, Rajasekaran S, Prokop JW. 2020. SARS-CoV-2 (COVID-19) structural and evolutionary dynamicome: Insights into functional evolution and human genomics. *J Biol Chem* **295**:11742–11753. doi:10.1074/jbc.ra120.014873

Gupta S, Singh AK, Kushwaha PP, Prajapati KS, Shuaib M, Senapati S, Kumar S. 2020. Identification of potential natural inhibitors of SARS-CoV2 main protease by molecular docking and simulation studies. *J Biomol Struct Dyn* 1–12. doi:10.1080/07391102.2020.1776157

Ho D, Wang P, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong P, Graham B, Mascola J, Chang J, Yin M, Sobieszczyk M, Kyratsous C, Shapiro L, Sheng Z, Nair M, Huang Y. 2021. Increased Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7 to Antibody Neutralization. *Res Sq*. doi:10.21203/rs.3.rs-155394/v1

Hoffmann M, Mösbauer K, Hofmann-Winkler H, Kaul A, Kleine-Weber H, Krüger N, Gassen NC, Müller MA, Drosten C, Pöhlmann S. 2020. Chloroquine does not inhibit infection of human lung cells with SARS-CoV-2. *Nature*. doi:10.1038/s41586-020-2575-3

Jeong GU, Song H, Yoon GY, Kim D, Kwon Y-C. 2020. Therapeutic Strategies Against COVID-19 and Structural Characterization of SARS-CoV-2: A Review. *Front Microbiol* **11**:1723. doi:10.3389/fmicb.2020.01723

Jogalekar MP, Veerabathini A, Gangadaran P. 2020. Novel 2019 coronavirus: Genome structure, clinical trials, and outstanding questions. *Exp Biol Med (Maywood)* **245**:964–969. doi:10.1177/1535370220920540

Kang S, Yang M, Hong Z, Zhang L, Huang Z, Chen X, He S, Zhou Ziliang, Zhou Zhechong, Chen Q, Yan Y, Zhang C, Shan H, Chen S. 2020. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B* **10**:1228–1238. doi:10.1016/j.apsb.2020.04.009

Kassab MA, Yu LL, Yu X. 2020. Targeting dePARylation for cancer therapy. *Cell Biosci* **10**:7. doi:10.1186/s13578-020-0375-y

Katsonis P, Lichtarge O. 2019. CAGI5: Objective performance assessments of predictions based on the Evolutionary Action equation. *Hum Mutat* **40**:1436–1454. doi:10.1002/humu.23873

Katsonis P, Lichtarge O. 2017. Objective assessment of the evolutionary action equation for the fitness effect of missense mutations across CAGI-blinded contests. *Hum Mutat* **38**:1072–1084. doi:10.1002/humu.23266

Katsonis P, Lichtarge O. 2014. A formal perturbation equation between genotype and phenotype determines the Evolutionary Action of protein-coding variations on fitness. *Genome Res* **24**:2050–2058. doi:10.1101/gr.176214.114

Kern DM, Sorum B, Hoel CM, Sridharan S, Remis JP, Toso DB, Brohawn SG. 2020. Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv Prepr Serv Biol*. doi:10.1101/2020.06.17.156554

Kim Y., Maltseva N, Jedrzejczak R, Endres M, Chang C, Godzik A, Michalska K, Joachimiak A, Center for Structural Genomics of Infectious Diseases (CSGID). 2020. RCSB PDB - 6WLC: Crystal Structure of NSP15 Endoribonuclease from SARS CoV-2 in the Complex with Uridine-5'-Monophosphate. doi:10.2210/pdb6WLC/pdb

Kim Youngchang, Wower J, Maltseva N, Chang C, Jedrzejczak R, Wilamowski M, Kang S, Nicolaescu V, Randall G, Michalska K, Joachimiak A. 2020. Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-CoV-2. *bioRxiv* 2020.06.26.173872. doi:10.1101/2020.06.26.173872

Kneller DW, Phillips G, O'Neill HM, Jedrzejczak R, Stols L, Langan P, Joachimiak A, Coates L, Kovalevsky A. 2020. Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat Commun* **11**. doi:10.1038/s41467-020-16954-7

Kraemer MUG, Yang C-H, Gutierrez B, Wu C-H, Klein B, Pigott DM, du Plessis L, Faria NR, Li R, Hanage WP, Brownstein JS, Layan M, Vespignani A, Tian H, Dye C, Pybus OG, Scarpino S V. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science (80-)* **368**:493–497. doi:10.1126/science.abb4218

Le Bert N, Tan AT, Kunasegaran K, Tham CYLL, Hafezi M, Chia A, Chng MHY, Lin M, Tan N, Linster M, Chia WN, Chen MI-C, Wang L-FF, Ooi EE, Kalimuddin S, Tambyah PA, Low JG-HH, Tan Y-JJ, Bertoletti A. 2020. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature*. doi:10.1038/s41586-020-2550-z

Lei J, Kusov Y, Hilgenfeld R. 2018. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res*. doi:10.1016/j.antiviral.2017.11.001

Li H, Liu S-M, Yu X-H, Tang S-L, Tang C-K. 2020. Coronavirus disease 2019 (COVID-19):

current status and future perspectives. *Int J Antimicrob Agents* **55**:105951. doi:10.1016/j.ijantimicag.2020.105951

Lichtarge O, Bourne HR, Cohen FE. 1996. An Evolutionary Trace Method Defines Binding Surfaces Common to Protein Families. *J Mol Biol* **257**:342–358. doi:10.1006/jmbi.1996.0167

Lichtarge O, Sowa ME, Philippi A. 2002. Evolutionary traces of functional surfaces along G protein signaling pathwayMethods in Enzymology. Academic Press Inc. pp. 536–556. doi:10.1016/S0076-6879(02)44739-8

Lin SM, Lin SC, Hsu JN, Chang CK, Chien CM, Wang YS, Wu HY, Jeng US, Kehn-Hall K, Hou MH. 2020. Structure-Based Stabilization of Non-native Protein-Protein Interactions of Coronavirus Nucleocapsid Proteins in Antiviral Drug Design. *J Med Chem* **63**:3131–3141. doi:10.1021/acs.jmedchem.9b01913

Lin SY, Liu CL, Chang YM, Zhao J, Perlman S, Hou MH. 2014. Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target. *J Med Chem* **57**:2247–2257. doi:10.1021/jm500089r

Littler DR, Gully BS, Colson RN, Rossjohn J. 2020. Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9. *iScience* **23**. doi:10.1016/j.isci.2020.101258

Liu B, Shi W, Yang Y. 2020. RCSB PDB - 6XQB: SARS-CoV-2 RdRp/RNA complex. doi:10.2210/pdb6XQB/pdb

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**:565–574. doi:10.1016/S0140-6736(20)30251-8

Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z. 2015. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc Natl Acad Sci U S A* **112**:9436–9441. doi:10.1073/pnas.1508686112

Madu IG, Roth SL, Belouzard S, Whittaker GR. 2009. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J Virol* **83**:7411–21. doi:10.1128/JVI.00079-09

Mateus J, Grifoni A, Tarke A, Sidney J, Ramirez SI, Dan JM, Burger ZC, Rawlings SA, Smith DM, Phillips E, Mallal S, Lammers M, Rubiro P, Quiambao L, Sutherland A, Yu ED, da Silva Antunes R, Greenbaum J, Frazier A, Markmann AJ, Premkumar L, de Silva A, Peters B, Crotty S, Sette A, Weiskopf D. 2020. Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science (80-)* eabd3871. doi:10.1126/science.abd3871

Meckiff BJ, Ramírez-Suástegui C, Fajardo V, Chee SJ, Kusnadi A, Simon H, Grifoni A, Pelosi E, Weiskopf D, Sette A, Ay F, Seumois G, Ottensmeier CH, Vijayanand P. 2020. Single-cell transcriptomic analysis of SARS-CoV-2 reactive CD4 + T cells. *bioRxiv Prepr Serv Biol*. doi:10.1101/2020.06.12.148916

Mihalek I, Reš I, Lichtarge O. 2007. Background frequencies for residue variability estimates: BLOSUM revisited. *BMC Bioinformatics* **8**. doi:10.1186/1471-2105-8-488

Mihalek I, Reš I, Lichtarge O. 2004. A Family of Evolution-Entropy Hybrid Methods for Ranking Protein Residues by Importance. *J Mol Biol* **336**:1265–1282. doi:10.1016/j.jmb.2003.12.078

Minasov G, Shuvalova L, Rosas-Lemus M, Kiryukhina O, Satchell KJF, Center for Structural Genomics of Infectious Diseases (CSGID). 2020a. RCSB PDB - 6WKQ: 1.98 Angstrom Resolution Crystal Structure of NSP16-NSP10 Heterodimer from SARS-CoV-2 in Complex with Sinefungin. doi:10.2210/pdb6WKQ/pdb

Minasov G, Shuvalova L, Rosas-Lemus M, Kiryukhina O, Wiersum G, Godzik A, Jaroszewski L, Stogios PJ, Skarina T, Satchell KJF, Center for Structural Genomics of Infectious Diseases (CSGID). 2020b. RCSB PDB - 6W4H: 1.80 Angstrom Resolution Crystal Structure of NSP16 - NSP10 Complex from SARS-CoV-2. doi:10.2210/pdb6W4H/pdb

Miyoshi-Akiyama T, Ishida I, Fukushi M, Yamaguchi K, Matsuoka Y, Ishihara T, Tsukahara M, Hatakeyama S, Itoh N, Morisawa A, Yoshinaka Y, Yamamoto N, Lianfeng Z, Chuan Q, Kirikae T, Sasazuki T. 2011. Fully Human Monoclonal Antibody Directed to Proteolytic Cleavage Site in Severe Acute Respiratory Syndrome (SARS) Coronavirus S Protein Neutralizes the Virus in a Rhesus Macaque SARS Model. *J Infect Dis* **203**:1574–1581. doi:10.1093/infdis/jir084

Mullard A. 2020. COVID-19 vaccine development pipeline gears up. *Lancet (London, England)* **395**:1751–1752. doi:10.1016/S0140-6736(20)31252-6

Nelson CA, Minasov G, Shuvalova L, Fremont DH, Center for Structural Genomics of Infectious Diseases (CSGID). 2020. RCSB PDB - 6W37: STRUCTURE OF THE SARS-CoV-2 ORF7A ENCODED ACCESSORY PROTEIN. doi:10.2210/pdb6W37/pdb

Newman JA, Yosaatmadja Y, Douangamath A, Arrowsmith CH, von Delft F, Edwards A, Bountra C, Gileadi O. 2020. RCSB PDB - 6ZSL: Crystal structure of the SARS-CoV-2 helicase at 1.94 Angstrom resolution. doi:10.2210/pdb6ZSL/pdb

Nolan S, Vignali M, Kaplan IM, Biotechnologies A, Svejnoha E, Craft T, Boland K, Pesesky M, Gittelman RM, Snyder TM, Gooley CJ, Semprini S, Istituto CC, Romagnolo S. 2020. A large-scale database of T-cell receptor beta (TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. Mark Klinger Adaptive Biotechnologies Jennifer N. Dines Adaptive Biotechnologies. doi:10.21203/rs.3.rs-51964/v1

Non-structural protein 4 (nsp4) | P0DTD1 PRO_0000449622 | Models. n.d. https://swissmodel.expasy.org/interactive/0cKKrV/models/01

Onrust R, Herzmark P, Chi P, Garcia PD, Lichtarge O, Kingsley C, Bourne HR. 1997. Receptor and βγ binding sites in the α subunit of the retinal G protein transducin. *Science (80-)* **275**:381–384. doi:10.1126/science.275.5298.381

Ortega JT, Serrano ML, Jastrzebska B. 2020. Class A G Protein-Coupled Receptor Antagonist Famotidine as a Therapeutic Alternative Against SARS-CoV2: An In Silico Analysis. *Biomolecules* **10**. doi:10.3390/biom10060954

Osipiuk J, Azizi S-A, Dvorkin S, Endres M, Jedrzejczak R, Jones KA, Kang S, Kathayat RS, Kim Y, Lisnyak VG, Maki SL, Nicolaescu V, Taylor CA, Tesar C, Zhang Y-A, Zhou Z, Randall G, Michalska K, Snyder SA, Dickinson BC, Joachimiak A. 2020. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *bioRxiv* 2020.08.06.240192. doi:10.1101/2020.08.06.240192

Owen CD, Lukacik P, Strain-Damerell CM, Douangamath A, Powell AJ, Fearon D, Brandao-Neto J, Crawshaw AD, Aragao D, Williams M, Flaig R, Hall DR, McAuley KE, Mazzorana M, Stuart DI, von Delft F, Walsh MA. 2020. RCSB PDB - 6YB7: SARS-CoV-2 main protease with unliganded active site (2019-nCoV, coronavirus disease 2019, COVID-19). doi:10.2210/pdb6YB7/pdb

Pancer K, Milewska A, Owczarek K, Dabrowska A, Kowalski M, Łabaj PP, Branicki W, Sanak M, Pyrc K. 2020. The SARS-CoV-2 ORF10 is not essential in vitro or in vivo in humans. *PLoS Pathog* **16**:e1008959. doi:10.1371/journal.ppat.1008959

Peng Y, Du N, Lei Y, Dorje S, Qi J, Luo T, Gao GF, Song H. 2020. Structures of the SARS -CoV-2 nucleocapsid and their perspectives for drug design . *EMBO J* **39**. doi:10.15252/embj.2020105938

Peterson SM, Pack TF, Wilkins AD, Urs NM, Urban DJ, Bass CE, Lichtarge O, Caron MG. 2015. Elucidation of G-protein and β-arrestin functional selectivity at the dopamine D2 receptor. *Proc Natl Acad Sci U S A* **112**:7097–102. doi:10.1073/pnas.1502742112

Pillaiyar T, Meenakshisundaram S, Manickam M. 2020. Recent discovery and development of inhibitors targeting coronaviruses. *Drug Discov Today* **25**:668–688. doi:10.1016/j.drudis.2020.01.015

Poh CM, Carissimo G, Wang B, Amrun SN, Lee CY-PP, Chee RS-LL, Fong S-WW, Yeo NK-WW, Lee W-HH, Torres-Ruesta A, Leo Y-SS, Chen MI-C, Tan S-YY, Chai LYA, Kalimuddin S, Kheng SSG, Thien S-YY, Young BE, Lye DC, Hanson BJ, Wang C-II, Renia L, Ng LFPP. 2020. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun* **11**:2806. doi:10.1038/s41467-020-16638-2

Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Pérez Marc G, Moreira ED, Zerbini C, Bailey R, Swanson KA, Roychoudhury S, Koury K, Li P, Kalina W V, Cooper D, Frenck RW, Hammitt LL, Türeci Ö, Nell H, Schaefer A, Ünal S, Tresnan DB, Mather S, Dormitzer PR, Şahin U, Jansen KU, Gruber WC, C4591001 Clinical Trial Group. 2020. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* **383**:2603–2615. doi:10.1056/NEJMoa2034577

Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, Connor T, Peacock T, Robertson DL, Erik V. 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563

Riva L, Yuan S, Yin X, Martin-Sancho L, Matsunaga N, Pache L, Burgstaller-Muehlbacher S, De Jesus PD, Teriete P, Hull M V., Chang MW, Chan JFW, Cao J, Poon VKM, Herbert KM, Cheng K, Nguyen TTH, Rubanov A, Pu Y, Nguyen C, Choi A, Rathnasinghe R, Schotsaert M, Miorin L, Dejosez M, Zwaka TP, Sit KY, Martinez-Sobrido L, Liu WC, White KM, Chapman ME, Lendy EK, Glynne RJ, Albrecht R, Ruppin E, Mesecar AD, Johnson JR, Benner C, Sun R, Schultz PG, Su AI, García-Sastre A, Chatterjee AK, Yuen KY, Chanda SK. 2020. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* 1–11. doi:10.1038/s41586-020-2577-1

Rodriguez GJ, Yao R, Lichtarge O, Wensel TG. 2010. Evolution-guided discovery and recoding of allosteric pathway specificity determinants in psychoactive bioamine receptors. *Proc Natl Acad Sci U S A* **107**:7787–92. doi:10.1073/pnas.0914877107

Rosas-Lemus M, Minasov G, Shuvalova L, Inniss NL, Kiryukhina O, Wiersum G, Kim Y, Jedrzejczak R, Enders M, Jaroszewski L, Godzik A, Joachimiak A, Satchell KJF. 2020. The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv* 2020.04.17.047498. doi:10.1101/2020.04.17.047498

Rut W, Lv Z, Zmudzinski M, Patchett S, Nayak D, Snipas SJ, El Oualid F, Huang TT, Bekes M, Drag M, Olsen SK. 2020. Activity profiling and structures of inhibitor-bound SARS-CoV-2-

PLpro protease provides a framework for anti-COVID-19 drug design. *bioRxiv Prepr Serv Biol*. doi:10.1101/2020.04.29.068890

Saikatendu KS, Joseph JS, Subramanian V, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn P. 2007. Ribonucleocapsid Formation of Severe Acute Respiratory Syndrome Coronavirus through Molecular Action of the N-Terminal Domain of N Protein. *J Virol* **81**:3913–3921. doi:10.1128/jvi.02236-06

Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. 2020. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. *JAMA* **323**:1824–1836. doi:10.1001/jama.2020.6019

Shenoy SK, Drake MT, Nelson CD, Houtz DA, Xiao K, Madabushi S, Reiter E, Premont RT, Lichtarge O, Lefkowitz RJ. 2006. β-arrestin-dependent, G protein-independent ERK1/2 activation by the β2 adrenergic receptor. *J Biol Chem* **281**:1261–1273. doi:10.1074/jbc.M506576200

Shrock E, Fujimura E, Kula T, Timms RT, Lee I-H, Leng Y, Robinson ML, Sie BM, Li MZ, Chen Y, Logue J, Zuiani A, McCulloch D, Lelis FJN, Henson S, Monaco DR, Travers M, Habibi S, Clarke WA, Caturegli P, Laeyendecker O, Piechocka-Trocha A, Li JZ, Khatri A, Chu HY, MGH COVID-19 Collection & Processing Team, Villani A-C, Kays K, Goldberg MB, Hacohen N, Filbin MR, Yu XG, Walker BD, Wesemann DR, Larman HB, Lederer JA, Elledge SJ. 2020. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* **370**. doi:10.1126/science.abd4250

Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **22**. doi:10.2807/1560-7917.ES.2017.22.13.30494

Sowa ME, He W, Slep KC, Kercher MA, Lichtarge O, Wensel TG. 2001. Prediction and confirmation of a site critical for effector regulation of RGS domain activity. *Nat Struct Biol* **8**:234–237. doi:10.1038/84974

Sowa ME, He W, Wensel TG, Lichtarge O. 2000. A regulator of G protein signaling interaction surface linked to effector specificity. *Proc Natl Acad Sci U S A* **97**:1483–1488. doi:10.1073/pnas.030409597

Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. 2020. QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics* **36**:1765–1771. doi:10.1093/bioinformatics/btz828

Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, Liu W, Bi Y, Gao GF. 2016. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol* **24**:490–502. doi:10.1016/j.tim.2016.03.003

Surya W, Li Y, Torres J. 2018. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim Biophys Acta - Biomembr* **1860**:1309–1317. doi:10.1016/j.bbamem.2018.02.017

Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao M, Korsman S, Davies M-A, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Pond SLK, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T. 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike

mutations in South Africa. *medRxiv* 2020.12.21.20248640. doi:10.1101/2020.12.21.20248640

Thoms M, Buschauer R, Ameismeier M, Koepke L, Denk T, Hirschenberger M, Kratzat H, Hayn M, Mackens-Kiani T, Cheng J, Straub JH, Stürzel CM, Fröhlich T, Berninghausen O, Becker T, Kirchhoff F, Sparrer KMJ, Beckmann R. 2020. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science (80-)* **369**:eabc8665. doi:10.1126/science.abc8665

Ulferts R, Ziebuhr J. 2011. Nidovirus ribonucleases: Structures and functions in viral replication. *RNA Biol* **8**:295–304. doi:10.4161/rna.8.2.15196

van Doremalen N, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham JN, Port JR, Avanzato VA, Bushmaker T, Flaxman A, Ulaszewska M, Feldmann F, Allen ER, Sharpe H, Schulz J, Holbrook M, Okumura A, Meade-White K, Pérez-Pérez L, Edwards NJ, Wright D, Bissett C, Gilbride C, Williamson BN, Rosenke R, Long D, Ishwarbhai A, Kailath R, Rose L, Morris S, Powers C, Lovaglio J, Hanley PW, Scott D, Saturday G, de Wit E, Gilbert SC, Munster VJ. 2020. ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature*. doi:10.1038/s41586-020-2608-y

Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, Angus B, Baillie VL, Barnabas SL, Bhorat QE, Bibi S, Briner C, Cicconi P, Collins AM, Colin-Jones R, Cutland CL, Darton TC, Dheda K, Duncan CJA, Emary KRW, Ewer KJ, Fairlie L, Faust SN, Feng S, Ferreira DM, Finn A, Goodman AL, Green CM, Green CA, Heath PT, Hill C, Hill H, Hirsch I, Hodgson SHC, Izu A, Jackson S, Jenkin D, Joe CCD, Kerridge S, Koen A, Kwatra G, Lazarus R, Lawrie AM, Lelliott A, Libri V, Lillie PJ, Mallory R, Mendes AVA, Milan EP, Minassian AM, McGregor A, Morrison H, Mujadidi YF, Nana A, O'Reilly PJ, Padayachee SD, Pittella A, Plested E, Pollock KM, Ramasamy MN, Rhead S, Schwarzbold A V., Singh N, Smith A, Song R, Snape MD, Sprinz E, Sutherland RK, Tarrant R, Thomson EC, Török ME, Toshner M, Turner DPJ, Vekemans J, Villafana TL, Watson MEE, Williams CJ, Douglas AD, Hill AVS, Lambe T, Gilbert SC, Pollard AJ, Aban M, Abayomi F, Abeyskera K, Aboagye J, Adam M, Adams K, Adamson J, Adelaja YA, Adewetan G, Adlou S, Ahmed K, Akhalwaya Y, Akhalwaya S, Alcock A, Ali A, Allen ER, Allen L, Almeida TCDSC, Alves MPS, Amorim F, Andritsou F, Anslow R, Appleby M, Arbe-Barnes EH, Ariaans MP, Arns B, Arruda L, Azi P, Azi L, Babbage G, Bailey C, Baker KF, Baker M, Baker N, Baker P, Baldwin L, Baleanu I, Bandeira D, Bara A, Barbosa MAS, Barker D, Barlow GD, Barnes E, Barr AS, Barrett JR, Barrett J, Bates L, Batten A, Beadon K, Beales E, Beckley R, Belij-Rammerstorfer S, Bell J, Bellamy D, Bellei N, Belton S, Berg A, Bermejo L, Berrie E, Berry L, Berzenyi D, Beveridge A, Bewley KR, Bexhell H, Bhikha S, Bhorat AE, Bhorat ZE, Bijker E, Birch G, Birch S, Bird A, Bird O, Bisnauthsing K, Bittaye M, Blackstone K, Blackwell L, Bletchly H, Blundell CL, Blundell SR, Bodalia P, Boettger BC, Bolam E, Boland E, Bormans D, Borthwick N, Bowring F, Boyd A, Bradley P, Brenner T, Brown P, Brown C, Brown-O'Sullivan C, Bruce S, Brunt E, Buchan R, Budd W, Bulbulia YA, Bull M, Burbage J, Burhan H, Burn A, Buttigieg KR, Byard N, Cabera Puig I, Calderon G, Calvert A, Camara S, Cao M, Cappuccini F, Cardoso JR, Carr M, Carroll MW, Carson-Stevens A, Carvalho Y de M, Carvalho JAM, Casey HR, Cashen P, Castro T, Castro LC, Cathie K, Cavey A, Cerbino-Neto J, Chadwick J, Chapman D, Charlton S, Chelysheva I, Chester O, Chita S, Cho JS, Cifuentes L, Clark E, Clark M, Clarke A, Clutterbuck EA, Collins SLK, Conlon CP, Connarty S, Coombes N, Cooper C, Cooper R, Cornelissen L, Corrah T, Cosgrove C, Cox T, Crocker WEM, Crosbie S, Cullen L, Cullen D, Cunha DRMF, Cunningham C, Cuthbertson FC, Da Guarda SNF, da Silva LP, Damratoski BE, Danos Z, Dantas MTDC, Darroch P, Datoo MS, Datta C, Davids M, Davies SL, Davies H, Davis E, Davis Judith, Davis John, De Nobrega MMD, De Oliveira Kalid LM, Dearlove D, Demissie T, Desai A, Di Marco S, Di Maso C, Dinelli MIS, Dinesh T, Docksey C, Dold C, Dong T, Donnellan FR, Dos Santos T, dos Santos TG, Dos Santos EP, Douglas N, Downing C, Drake J, Drake-Brockman R, Driver K, Drury R, Dunachie SJ, Durham BS,

Dutra L, Easom NJW, van Eck S, Edwards M, Edwards NJ, El Muhanna OM, Elias SC, Elmore M, English M, Esmail A, Essack YM, Farmer E, Farooq M, Farrar M, Farrugia L, Faulkner B, Fedosyuk S, Felle S, Ferreira Da Silva C, Field S, Fisher R, Flaxman A, Fletcher J, Fofie H, Fok H, Ford KJ, Fowler J, Fraiman PHA, Francis E, Franco MM, Frater J, Freire MSM, Fry SH, Fudge S, Furze J, Fuskova M, Galian-Rubio P, Galiza E, Garlant H, Gavrila M, Geddes A, Gibbons KA, Gilbride C, Gill H, Glynn S, Godwin K, Gokani K, Goldoni UC, Goncalves M, Gonzalez IGS, Goodwin J, Goondiwala A, Gordon-Quayle K, Gorini G, Grab J, Gracie L, Greenland M, Greenwood N, Greffrath J, Groenewald MM, Grossi L, Gupta G, Hackett M, Hallis B, Hamaluba M, Hamilton E, Hamlyn J, Hammersley D, Hanrath AT, Hanumunthadu B, Harris SA, Harris C, Harris T, Harrison TD, Harrison D, Hart TC, Hartnell B, Hassan S, Haughney J, Hawkins S, Hay J, Head I, Henry J, Hermosin Herrera M, Hettle DB, Hill J, Hodges G, Horne E, Hou MM, Houlihan C, Howe E, Howell N, Humphreys J, Humphries HE, Hurley K, Huson C, Hyder-Wright A, Hyams C, Ikram S, Ishwarbhai A, Ivan M, Iveson P, Iyer V, Jackson F, De Jager J, Jaumdally S, Jeffers H, Jesudason N, Jones B, Jones K, Jones E, Jones C, Jorge MR, Jose A, Joshi A, Júnior EAMS, Kadziola J, Kailath R, Kana F, Karampatsas K, Kasanyinga M, Keen J, Kelly EJ, Kelly DM, Kelly D, Kelly S, Kerr D, Kfouri R de Á, Khan L, Khozoee B, Kidd S, Killen A, Kinch J, Kinch P, King LDW, King TB, Kingham L, Klenerman P, Knapper F, Knight JC, Knott D, Koleva S, Lang M, Lang G, Larkworthy CW, Larwood JPJ, Law R, Lazarus EM, Leach A, Lees EA, Lemm NM, Lessa A, Leung S, Li Y, Lias AM, Liatsikos K, Linder A, Lipworth S, Liu S, Liu X, Lloyd A, Lloyd S, Loew L, Lopez Ramon R, Lora L, Lowthorpe V, Luz K, MacDonald JC, MacGregor G, Madhavan M, Mainwaring DO, Makambwa E, Makinson R, Malahleha M, Malamatsho R, Mallett G, Mansatta K, Maoko T, Mapetla K, Marchevsky NG, Marinou S, Marlow E, Marques GN, Marriott P, Marshall RP, Marshall JL, Martins FJ, Masenya M, Masilela M, Masters SK, Mathew M, Matlebjane H, Matshidiso K, Mazur O, Mazzella A, McCaughan H, McEwan J, McGlashan J, McInroy L, McIntyre Z, McLenaghan D, McRobert N, McSwiggan S, Megson C, Mehdipour S, Meijs W, Mendonça RNÁ, Mentzer AJ, Mirtorabi N, Mitton C, Mnyakeni S, Moghaddas F, Molapo K, Moloi M, Moore M, Moraes-Pinto MI, Moran M, Morey E, Morgans R, Morris Susan, Morris Sheila, Morris HC, Morselli F, Morshead G, Morter R, Mottal L, Moultrie A, Moya N, Mpelembue M, Msomi S, Mugodi Y, Mukhopadhyay E, Muller J, Munro A, Munro C, Murphy S, Mweu P, Myasaki CH, Naik G, Naker K, Nastouli E, Nazir A, Ndlovu B, Neffa F, Njenga C, Noal H, Noé A, Novaes G, Nugent FL, Nunes G, O'Brien K, O'Connor D, Odam M, Oelofse S, Oguti B, Olchawski V, Oldfield NJ, Oliveira MG, Oliveira C, Oosthuizen A, O'Reilly P, Osborne P, Owen DRJ, Owen L, Owens D, Owino N, Pacurar M, Paiva BVB, Palhares EMF, Palmer S, Parkinson S, Parracho HMRT, Parsons K, Patel D, Patel B, Patel F, Patel K, Patrick-Smith M, Payne RO, Peng Y, Penn EJ, Pennington A, Peralta Alvarez MP, Perring J, Perry N, Perumal R, Petkar S, Philip T, Phillips DJ, Phillips J, Phohu MK, Pickup L, Pieterse S, Piper J, Pipini D, Plank M, Du Plessis J, Pollard S, Pooley J, Pooran A, Poulton I, Powers C, Presa FB, Price DA, Price V, Primeira M, Proud PC, Provstgaard-Morys S, Pueschel S, Pulido D, Quaid S, Rabara R, Radford A, Radia K, Rajapaska D, Rajeswaran T, Ramos ASF, Ramos Lopez F, Rampling T, Rand J, Ratcliffe H, Rawlinson T, Rea D, Rees B, Reiné J, Resuello-Dauti M, Reyes Pabon E, Ribiero CM, Ricamara M, Richter A, Ritchie N, Ritchie AJ, Robbins AJ, Roberts H, Robinson RE, Robinson H, Rocchetti TT, Rocha BP, Roche S, Rollier C, Rose L, Ross Russell AL, Rossouw L, Royal S, Rudiansyah I, Ruiz S, Saich S, Sala C, Sale J, Salman AM, Salvador N, Salvador S, Sampaio M, Samson AD, Sanchez-Gonzalez A, Sanders H, Sanders K, Santos E, Santos Guerra MFS, Satti I, Saunders JE, Saunders C, Sayed A, Schim van der Loeff I, Schmid AB, Schofield E, Screaton G, Seddiqi S, Segireddy RR, Senger R, Serrano S, Shah R, Shaik I, Sharpe HE, Sharrocks K, Shaw R, Shea A, Shepherd A, Shepherd JG, Shiham F, Sidhom E, Silk SE, da Silva Moraes AC, Silva-Junior G, Silva-Reyes L, Silveira AD, Silveira MBV, Sinha J, Skelly DT, Smith DC, Smith N, Smith HE, Smith DJ, Smith CC, Soares A, Soares T, Solórzano C, Sorio GL, Sorley K, Sosa-Rodriguez T, Souza CMCDL, Souza BSDF, Souza AR, Spencer AJ, Spina F, Spoors L, Stafford L, Stamford I, Starinskij I, Stein R, Steven J, Stockdale L, Stockwell L

V., Strickland LH, Stuart AC, Sturdy A, Sutton N, Szigeti A, Tahiri-Alaoui A, Tanner R, Taoushanis C, Tarr AW, Taylor K, Taylor U, Taylor IJ, Taylor J, te Water Naude R, Themistocleous Y, Themistocleous A, Thomas M, Thomas K, Thomas TM, Thombrayil A, Thompson F, Thompson Amber, Thompson K, Thompson Ameeka, Thomson J, Thornton-Jones V, Tighe PJ, Tinoco LA, Tiongson G, Tladinyane B, Tomasicchio M, Tomic A, Tonks S, Towner J, Tran N, Tree J, Trillana G, Trinham C, Trivett R, Truby A, Tsheko BL, Turabi A, Turner R, Turner C, Ulaszewska M, Underwood BR, Varughese R, Verbart D, Verheul M, Vichos I, Vieira T, Waddington CS, Walker L, Wallis E, Wand M, Warbick D, Wardell T, Warimwe G, Warren SC, Watkins B, Watson E, Webb S, Webb-Bridges A, Webster A, Welch J, Wells J, West A, White C, White R, Williams P, Williams RL, Winslow R, Woodyer M, Worth AT, Wright D, Wroblewska M, Yao A, Zimmer R, Zizi D, Zuidewind P. 2021. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* **397**:99–111. doi:10.1016/S0140-6736(20)32661-1

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**:W296–W303. doi:10.1093/nar/gky427

White KM, Rosales R, Yildiz S, Kehrer T, Miorin L, Moreno E, Jangra S, Uccellini MB, Rathnasinghe R, Coughlan L, Martinez-Romero C, Batra J, Rojc A, Bouhaddou M, Fabius JM, Obernier K, Dejosez M, Guillén MJ, Losada A, Avilés P, Schotsaert M, Zwaka T, Vignuzzi M, Shokat KM, Krogan NJ, García-Sastre A. 2021. Plitidepsin has potent preclinical efficacy against SARS-CoV-2 by targeting the host protein eEF1A. *Science*. doi:10.1126/science.abf4058

Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Lambson BE, Vermeulen M, van den Berg K, Rossouw T, Boswell M, Ueckermann V, Meiring S, von Gottberg A, Cohen C, Morris L, Bhiman JN, Moore PL. 2021. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *bioRxiv Prepr Serv Biol*. doi:10.1101/2021.01.18.427166

Wilkins AD, Lua R, Erdin S, Ward RM, Lichtarge O. 2010. Sequence and structure continuity of evolutionary importance improves protein functional site discovery and annotation. *Protein Sci* **19**:1296–1311. doi:10.1002/pro.406

Wilkins AD, Venner E, Marciano DC, Erdin S, Atri B, Lua RC, Lichtarge O. 2013. Accounting for epistatic interactions improves the functional analysis of protein structures. *Bioinformatics* **29**:2714–2721. doi:10.1093/bioinformatics/btt489

Woo H, Park S-J, Choi YK, Park T, Tanveer M, Cao Y, Kern NR, Lee J, Yeom MS, Croll TI, Seok C, Im W. 2020. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *J Phys Chem B* **124**:7128–7137. doi:10.1021/acs.jpcb.0c04553

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, Yuan M-L, Zhang Y-L, Dai F-H, Liu Y, Wang Q-M, Zheng J-J, Xu L, Holmes EC, Zhang Y-Z. 2020. A new coronavirus associated with human respiratory disease in China. *Nature* **579**:265–269. doi:10.1038/s41586-020-2008-3

Wu K, Werner AP, Moliva JI, Koch M, Choi A, Stewart-Jones GBE, Bennett H, Boyoglu-Barnum S, Shi W, Graham BS, Carfi A, Corbett KS, Seder RA, Edwards DK. 2021. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. *bioRxiv Prepr Serv Biol*. doi:10.1101/2021.01.25.427948

Yin W, Mao C, Luan X, Shen D-DD, Shen Q, Su H, Wang X, Zhou F, Zhao W, Gao M, Chang S, Xie Y-CC, Tian G, Jiang HWHH-W, Tao S-CC, Shen J, Jiang Y, Jiang HWHH-W, Xu Y, Zhang S, Zhang Y, Xu HE. 2020. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**:1499–1504. doi:10.1126/science.abc1560

Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, Wu Q, Merler S, Viboud C, Vespignani A, Ajelli M, Yu H. 2020. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368**:1481–1486. doi:10.1126/science.abb8001

Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R. 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **412**:409–412. doi:10.1126/science.abb3405

Zhao W-M, Song S-H, Chen M-L, Zou D, Ma L-N, Ma Y-K, Li R-J, Hao L-L, Li C-P, Tian D-M, Tang B-X, Wang Y-Q, Zhu J-W, Chen H-X, Zhang Z, Xue Y-B, Bao Y-M. 2020. The 2019 novel coronavirus resource. *Yi chuan = Hered* **42**:212–221. doi:10.16288/j.yczz.20-030

Zinzula L, Basquin J, Nagy I, Bracher A. 2020. RCSB PDB - 6ZCO: Crystal Structure of C-terminal Dimerization Domain of Nucleocapsid Phosphoprotein from SARS-CoV-2, crystal form II. doi:10.2210/pdb6ZCO/pdb

Figures

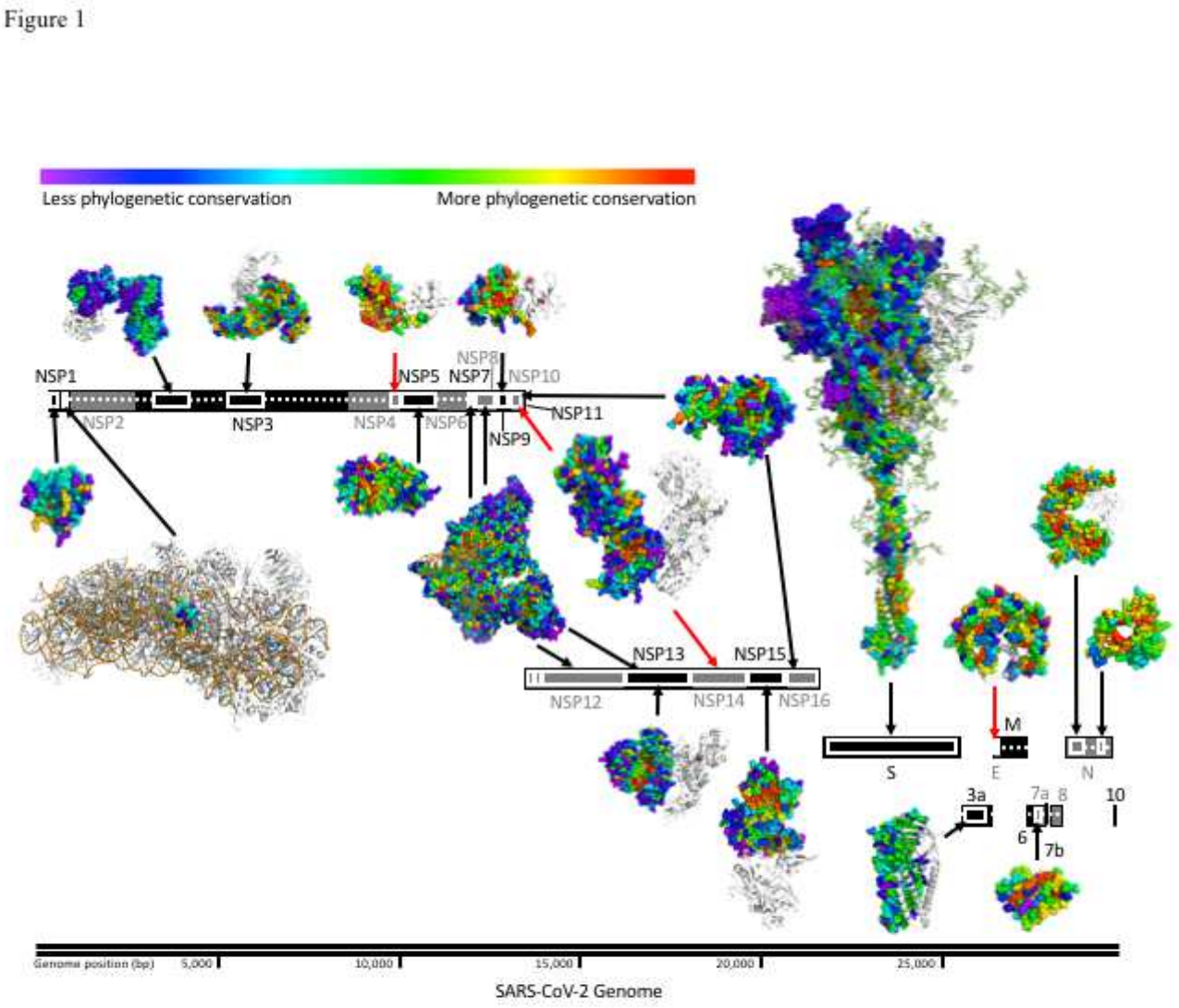

Figure 1

Structural and sequence information permits identification of evolutionarily important sites in SARS-CoV-2. Linear representation of SARS-CoV-2 proteome with structurally determined regions highlighted (white boxes) and corresponding structures' residues colored by Evolutionary Trace rank. Black arrows connect SARS-CoV-2 structures to their corresponding gene, red arrows indicate that only structures of homologous proteins are available. Host ribosomal proteins in the NSP1 complex structure are shown in

white. For multimeric structures, one monomer is also shown in white. Structures shown include: 7k7p (Clark et al., 2020) (NSP1), 6zlw (Thoms et al., 2020) (NSP1 C-term), 6woj (Alhammad et al., 2020) (NSP3), 6w9c (Osipiuk et al., 2020) (NSP3), ExPasy NSP4 model 01 (Bienert et al., 2017; "Non-structural protein 4 (nsp4) | P0DTD1 PRO_0000449622 | Models," n.d.; Studer et al., 2020; Waterhouse et al., 2018) (NSP4), 6yb7 (Owen et al., 2020) (NSP5), 6wxd (Littler et al., 2020) (NSP9), 6xez (Chen et al., 2020) (NSP7, 8 , 12, and 13), 6zsl (Newman et al., 2020) (NSP13), 5c8s (Ma et al., 2015) (NSP10 and 14), 6wlc (Y. Kim et al., 2020) (NSP15), 6w4h (Minasov et al., 2020b) (NSP10 and NSP16), 6vsb_1_1_1 (Woo et al., 2020) (S), 6xdc (Kern et al., 2020) (ORF3a), 5x29 (Surya et al., 2018) (E), 6w37 (Nelson et al., 2020) (ORF7a), 6vyo (Chang et al., 2020) (N), and 6zco (Zinzula et al., 2020) (N).

Figure 2

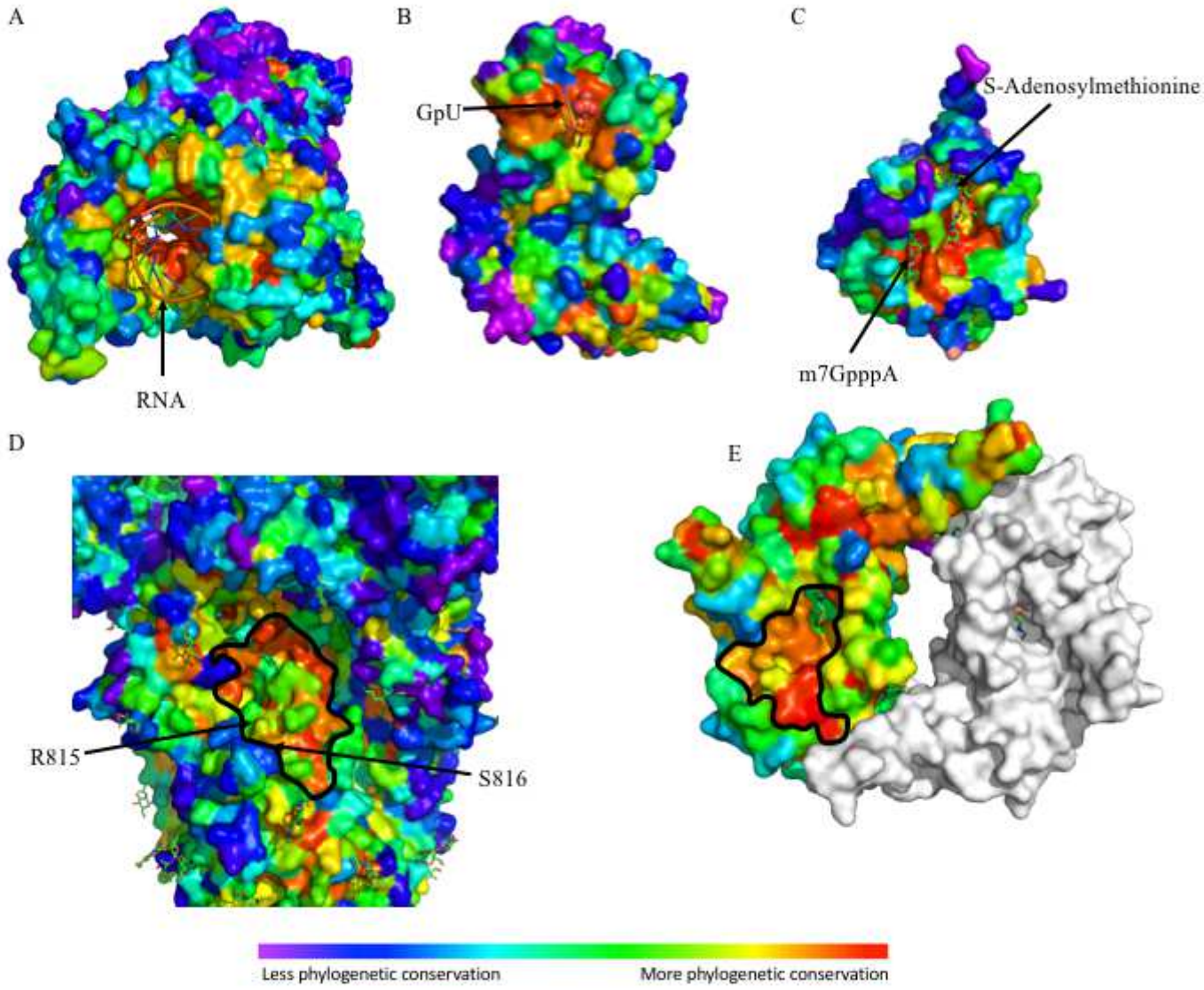

Figure 2

Top ET ranking residues overlap with known functional sites. ET recovers (A) the RNA binding site of NSP12 (RNA dependent RNA polymerase, pdb:6xez), (B) the active site of NSP15 (uridine-specific endoribonuclease, pdb:6x1b), (C) the substrates binding sites of NSP16 (RNA-cap methyltransferase, pdb:6wvn). ET also recovers a key functional (D) S2’ protease cleavage site of S (key residues: R815 and S816, pdb:6vsb), and predicts (E) a site associated with the putative RNA binding site of N (pdb:6vyo). The S cleavage site between R815 and S816 is labeled and the putative sites of panels D and E are highlighted with a black outline. The site in panel D is a high priority target as it was found that residues 815-825, which overlap this site, comprise the most frequently recognized epitope among naïve and COVID-19 patients (Shrock et al., 2020).

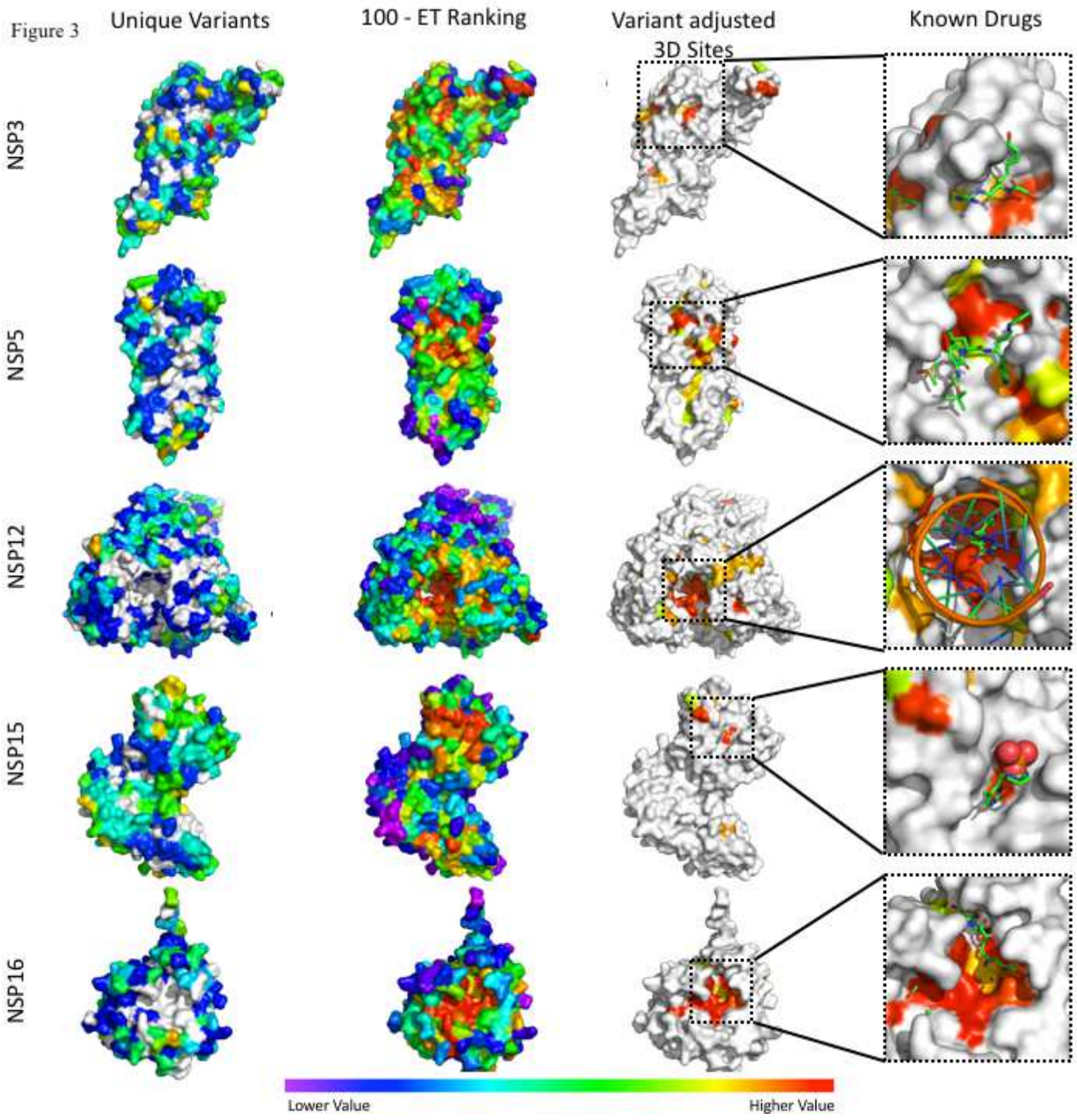

Figure 3

Identification of variant adjusted 3D sites (5Å) and their colocalization with known drug binding sites. Variant adjusted 3D sites for NSP3 (6w9c(Osipiuk et al., 2020)), NSP5 (6yb7(Owen et al., 2020)), NSP12 (7bv1(Yin et al., 2020)), NSP15 (6wlc(Youngchang Kim et al., 2020)), and NSP16 (6w4h(Minasov et al., 2020b)) were identified as clusters of surface residues with low ET ranks and a lack of mutations in the current outbreak. In the known drugs panels, variant adjusted 3D sites were identified using apo form structures, then mapped to the co-structures of NSP3 with peptide inhibitor vir251 (PDB:6wx4(Rut et al., 2020)), NSP5 with potential drug 13b (PDB:6y2f(L. Zhang et al., 2020)), NSP12 with drug remdesivir

(7bv2(Yin et al., 2020)), NSP15 in complex with potential drug tipiracil (PDB:6wxc(Youngchang Kim et al., 2020)), and NSP16 with sinefungin (PDB:6wkq(Minasov et al., 2020a)). For structures in the “Unique Variants” column “Lower Values” in the color scale correspond to fewer variants, while “Higher Values” correspond to more variants and white residues have no reported variants in the analyzed SARS-CoV-2 strains. For the “100 – ET Ranking”, “Variant Adjusted 3D Sites”, and “Known Drugs” columns, “Lower Values” correspond to less phylogenetic conservation while “Higher Values” correspond to more phylogenetic conservation.

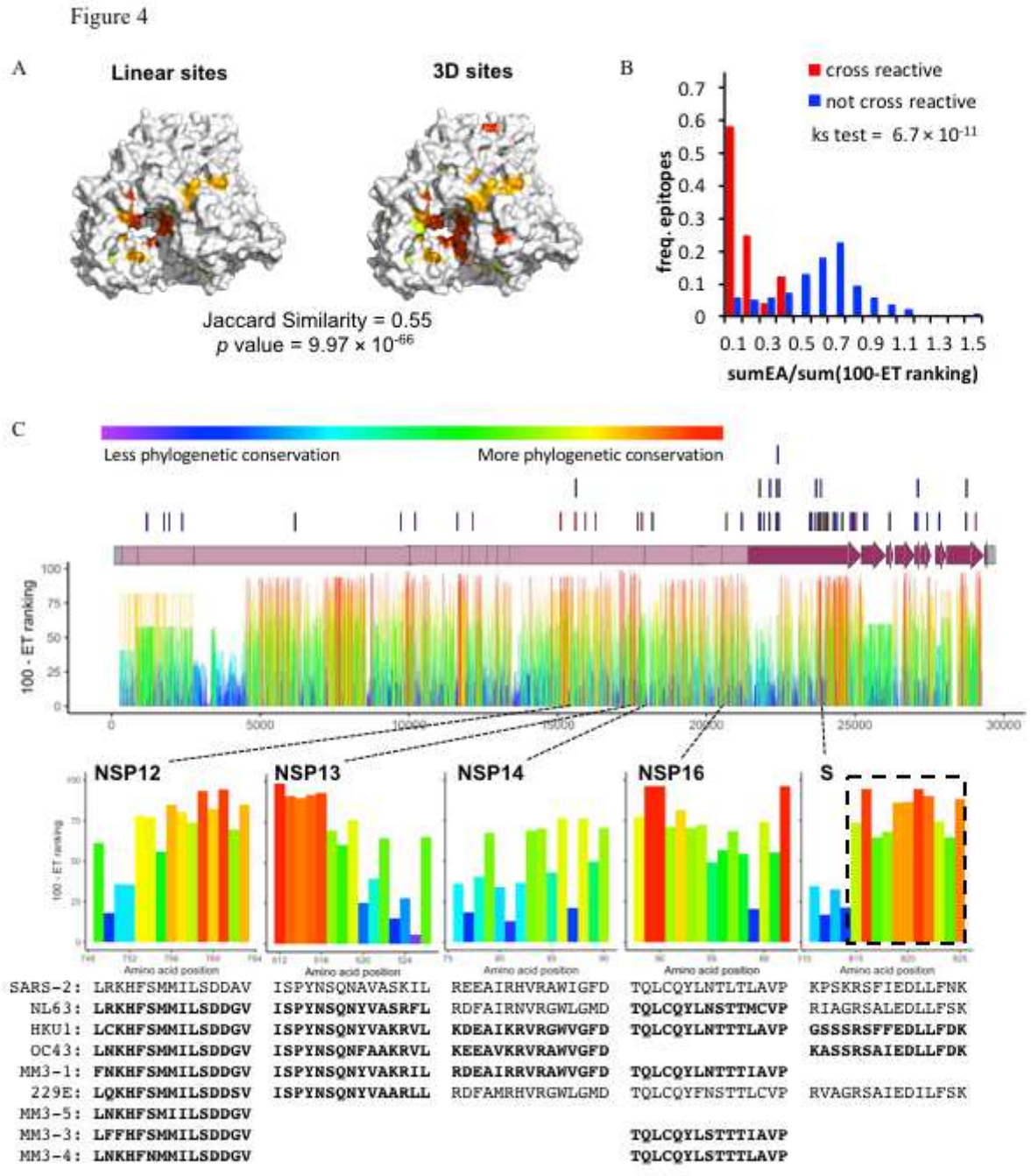

Figure 4

Identification of variant adjusted linear sites with Evolutionary Trace. A) Mapping of variant adjusted linear and structural sites on the surface of NSP12 (7bv1) with Jaccard Similarity value and Fisher's Exact test p value indicated. B) Relative frequency distributions of sumEA/sum(100-ET ranking) for T-cell epitopes shown to either be cross reactive (red) or not (blue). The sumEA/sum(100-ET ranking) metric predicts the functional impact of variants (EA) relative to the overall Evolutionary Trace rankings in the epitope. A Kolmogorov-Smirnov test (ks test) shows a significant difference in the distributions. C) T-cell epitopes reported in Mateus, et al. (Mateus et al., 2020) are shown above the SARS-CoV-2 genome (lines) and ET rankings within each protein are shown below. Shown are five SARS-CoV-2 epitopes (NSP12, NSP13, NSP14, NSP16 or S) that are predicted to cross react with the indicated common human coronavirus epitopes (bold text). Closely related coronavirus epitopes that did not meet our stringent threshold are also shown (normal text). The dashed box highlights the 11 amino acid stretch subsequently shown to be the most cross-reactive Spike protein epitope among naïve and COVID-19 patients (Shrock et al., 2020).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ETCoV2R2SIv4.pdf