OXFORD

Briefings in Bioinformatics, 2024, **25(3)**, bbae183
https://doi.org/10.1093/bib/bbae183
Problem Solving Protocol

CovEpiAb: a comprehensive database and analysis resource for immune epitopes and antibodies of human coronaviruses

Xue Zhang, JingCheng Wu, Yuanyuan Luo, Yilin Wang, Yujie Wu, Xiaobin Xu, Yufang Zhang, Ruiying Kong, Ying Chi, Yisheng Sun, Shuqing Chen, Qiaojun He, Feng Zhu and Zhan Zhou

Corresponding authors. Feng Zhu, State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University, #866 Yuhangtang Road, Hangzhou 310058, China. Tel.: +86-571-88208444; Fax: +86-571-88208444; E-mail: zhufeng@zju.edu.cn; Zhan Zhou, State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University, #866 Yuhangtang Road, Hangzhou 310058, China. Tel.: +86-571-88208410; Fax: +86-571-88208410; E-mail: zhanzhou@zju.edu.cn

Abstract

Coronaviruses have threatened humans repeatedly, especially COVID-19 caused by SARS-CoV-2, which has posed a substantial threat to global public health. SARS-CoV-2 continuously evolves through random mutation, resulting in a significant decrease in the efficacy of existing vaccines and neutralizing antibody drugs. It is critical to assess immune escape caused by viral mutations and develop broad-spectrum vaccines and neutralizing antibodies targeting conserved epitopes. Thus, we constructed CovEpiAb, a comprehensive database and analysis resource of human coronavirus (HCoVs) immune epitopes and antibodies. CovEpiAb contains information on over 60 000 experimentally validated epitopes and over 12 000 antibodies for HCoVs and SARS-CoV-2 variants. The database is unique in (1) classifying and annotating cross-reactive epitopes from different viruses and variants; (2) providing molecular and experimental interaction profiles of antibodies, including structure-based binding sites and around 70 000 data on binding affinity and neutralizing activity; (3) providing virological characteristics of current and past circulating SARS-CoV-2 variants and *in vitro* activity of various therapeutics; and (4) offering site-level annotations of key functional features, including antibody binding, immunological epitopes, SARS-CoV-2 mutations and conservation across HCoVs. In addition, we developed an integrated pipeline for epitope prediction named COVEP, which is available from the webpage of CovEpiAb. CovEpiAb is freely accessible at https://pgx.zju.edu.cn/covepiab/.

Keywords: human coronaviruses (HCoVs); SARS-CoV-2 variants; immune epitope; antibodies; vaccines

Xue Zhang is a PhD candidate at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University.
JingCheng Wu is a postdoctoral associate at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University.
Yuanyuan Luo is a graduate student at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University.
Yilin Wang is an undergraduate student at the College of Pharmaceutical Sciences, Zhejiang University.
Yujie Wu is an undergraduate student at the College of Pharmaceutical Sciences, Zhejiang University.
Xiaobin Xu is a graduate student at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University.
Yufang Zhang is an undergraduate student at the College of Pharmaceutical Sciences, Zhejiang University.
Ruiying Kong is a graduate student at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University.
Ying Chi is an associate professor at ZJU-UoE Institute, Zhejiang University.
Yisheng Sun is a researcher at the Prevention and Control of Infectious Disease of Zhejiang Province, China.
Shuqing Chen is a professor at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University.
Qiaojun He is a professor at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University. He is also a professor at the Zhejiang University Innovation Institute for Artificial Intelligence in Medicine, Engineering Research Center of Innovative Anticancer Drugs, Ministry of Education.
Feng Zhu is a professor at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University. He is also a professor at the Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare.
Zhan Zhou is an associate professor at the State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University and an associate professor at the Zhejiang University Innovation Institute for Artificial Intelligence in Medicine, Engineering Research Center of Innovative Anticancer Drugs, Ministry of Education. He is also an adjunct professor at Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, and The Fourth Affiliated Hospital, Zhejiang University School of Medicine.
Received: January 3, 2024. **Revised:** February 24, 2024. **Accepted:** April 8, 2024

INTRODUCTION

The COVID-19 global pandemic caused by SARS-CoV-2 had serious consequences for the global economy and human health [1, 2]. SARS-CoV-2 continuously evolves through random mutations, with new variants potentially enhancing transmissibility, virulence and the ability to evade adaptive immune responses. Evidence suggests that certain SARS-CoV-2 variants exhibit reduced sensitivity to plasma from individuals with previous infections or vaccinations, as well as some anti-SARS-CoV-2 antibodies [3–9]. Furthermore, monoclonal antibody (mAb) products previously authorized for COVID-19 treatment are currently ineffective against the circulating Omicron variants [10, 11]. The efficacy of vaccines against variants carrying these mutations is also compromised [7–9]. These issues underscore the critical importance of monitoring and assessing the immune escape of variants and developing broad and effective prevention and therapeutic strategies.

Human coronaviruses (HCoVs), including SARS-CoV, MERS-CoV and four other common coronaviruses, namely, HCoV-229E, HKU1, OC43 and NL63, have previously caused outbreaks or respiratory symptoms [12–16]. They share genomic similarity and evolutionary conservation with SARS-CoV-2, with important implications for vaccine development and pandemic prevention. Studies have shown that individuals previously exposed to other HCoVs can develop protection against SARS-CoV-2, indicating the potential cross-protection of pre-existing T cells in response to stimulation of conserved antigenic regions [17–22]. Broadly neutralizing antibodies against SARS-CoV-2 variants commonly exhibit recognition toward conserved regions in HCoVs [23, 24], which has supported the development of immunization strategies based on evolutionary conservation. Therefore, it is crucial to establish a database encompassing immune epitopes and antibodies targeting HCoVs and SARS-CoV-2 variants.

To date, various databases have been established to explore SARS-CoV-2 and related coronaviruses. Public repositories such as GISAID [25], NCBI [26] and RCoV19 [27] play a crucial role in storing and managing sequence data. Several web tools and databases such as Nextstrain [28], cov-lineages [29], CoVizu [30] and Outbreak.info [31] track the evolution and transmission dynamics of SARS-CoV-2 using public sequences. Databases like ViruSurf [32], GESS [33], CoV-RDB [34], VarEPS [35] and CoV3D [36] annotate the sequence, structure and function of the viral genome. Some resources such as COVIEdb [37], CoV-AbDab [38] and Ab-CoV [39] only provide lists of epitopes and antibodies, while others such as PAGER-CoV [40], CovInter [41], DockCoV2 [42] and COVID19 Drug Repository [43] describe virus–host interactions, drug targets and drug repurposing. However, no database provides immunoinformatic resources specifically for HCoV evolutionary conservation. To address this gap, we performed comprehensive bioinformatics analysis and annotation of immune epitopes and antibody interactions associated with HCoVs and SARS-CoV-2 variants.

In this study, we constructed CovEpiAb, a functional database and analysis resource for immune epitopes and antibodies of HCoVs (Figure 1). We collected a large collection of experimentally validated epitope and antibody data, including 66 210 B-cell epitopes, 3209 T-cell epitopes and 12 613 antibodies with detailed annotations. For epitopes, we examined and classified cross-reactive epitopes from different coronaviruses and variants. For antibodies, we gathered approximately 70 000 experimental data points that delineate the interaction profiles of antibody binding affinity and neutralizing activity. Additionally, we elucidated binding sites through the analysis of antigen–antibody complex structures. For current and past circulating SARS-CoV-2 variants, the database provides virological characteristics and *in vitro* activity data against advanced therapeutics, including vaccines, neutralizing antibodies and convalescent plasma. At the site level, we depicted the functional annotation of immunogenic epitopes, antibody binding sites and SARS-CoV-2 spike mutations. Finally, the database offers an epitope prediction and analysis resource, an integrated online pipeline named COVEP, for the prediction of B-cell and T-cell epitopes. Overall, CovEpiAb provides an important resource for SARS-CoV-2-related research and facilitates the development of broad-spectrum effective vaccines and neutralizing antibodies against HCoVs.

METHODS

Epitope collection and processing

We retrieved data from the Immune Epitope Database and Analysis Resource (IEDB) [44] and searched PubMed by keywords such as 'identify SARS-CoV-2 epitopes', 'identify SARS-CoV epitopes', 'identify MERS-CoV epitopes' and 'identify human coronavirus epitopes'. After merging and preprocessing, we identified a total of 66 210 B-cell epitopes and 3209 T-cell epitopes with 597 references. All the collected epitopes were experimentally confirmed positive (reactive) epitopes in the literature. Epitope identification primarily relied on qualitative measurements, with a few using quantitative measurements such as dissociation constant (KD), but there were no standardized thresholds in the literature, and the experimental assays were provided in the epitope detail annotations.

To reveal potential mutation sites affecting immunogenicity, we determined the overlap between SARS-CoV-2 epitopes and spike mutations of variants by computational matching and plotted heatmaps. To determine the conservation of epitopes among HCoVs, we used MAFFT [45] to conduct multiple sequence alignments of five homologous proteins of HCoVs, including spike (S), envelope (E), membrane (M), nucleoprotein (N) and open reading frame 1ab (ORF1ab). The 'conserv' method in Bio3D [46] was used to calculate the site conservation score and pyMSAviz [47] was used to visualize the sequence alignment. Residues with a score above 0.7 were considered conserved. Epitope conservation scores were computed by averaging the scores of all sites within the peptide. Conservation sequences were defined as contexts of conserved sites with no more than three gaps between them and longer than seven residues. Then, at the site level, we depicted the functional annotations of all sites within the epitope, indicating whether the site functions as an ACE-2 receptor binding site, has undergone mutations, is an mAb resistance mutation site and its conservation score across HCoVs.

Cross-reactive epitopes definition and classification

Identification of positive epitopes relied on experimental methods to assess reactivity between T/B cells and candidate epitope peptides. Samples were obtained from healthy, vaccinated, infected or environmentally exposed individuals with no evidence of disease. For T-cell epitopes, *in vitro* stimulation may be used to reactivate responsive T cells. *In vivo* and *in vitro* stimulatory immunogens include viral organisms, proteins and peptides (Figure S1). Epitopes identified from samples without immunogen stimulation were termed 'Undetermined', while those identified from environmentally exposed individuals without disease were termed 'Enviro-exposed' epitopes.

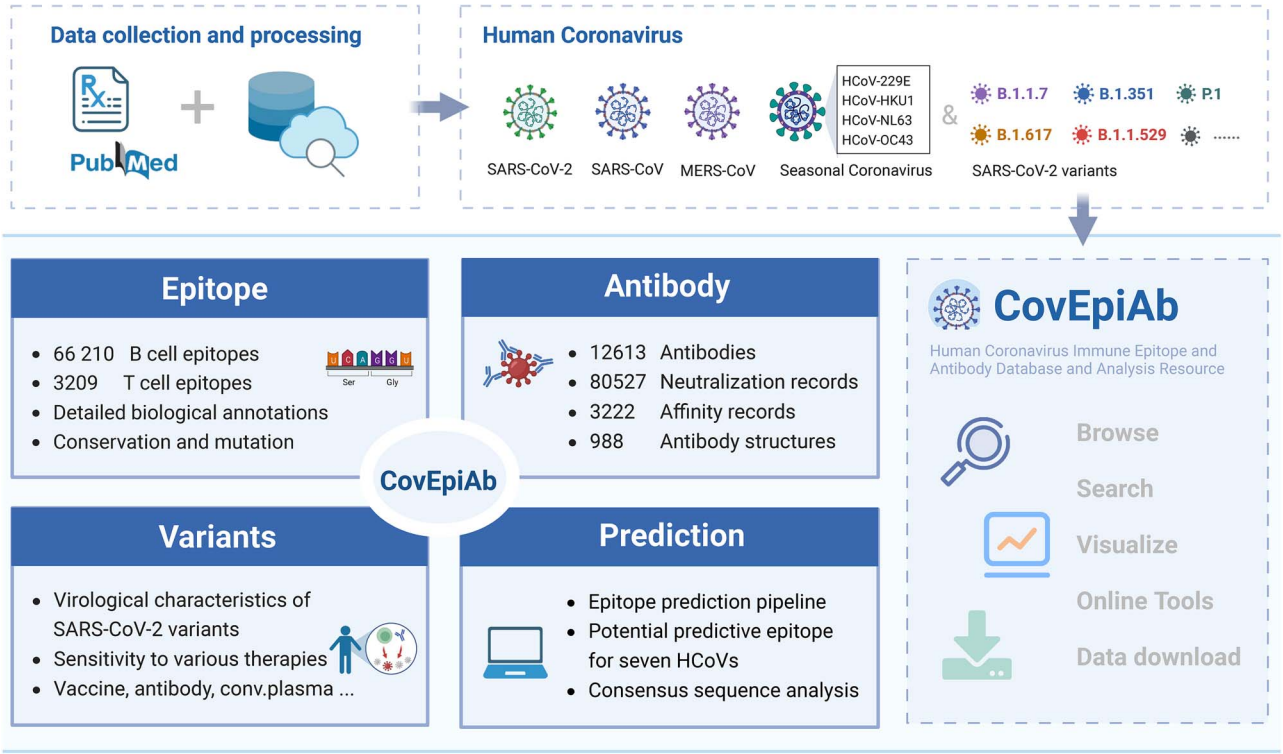

Figure 1. Workflow and overview of CovEpiAb database. Data were collected from public databases and literature and then were preprocessed. These curated resources include immune epitopes and antibodies of HCoVs, virological characteristics of SARS-CoV-2 variants and *in vitro* activity data of various therapeutics. Furthermore, to facilitate epitope prediction and analysis, an online pipeline named COVEP was developed, which integrates existing B-cell and T-cell epitope prediction tools. Finally, the CovEpiAb database was constructed with MySQL and Django tools, incorporating browsing, searching, visualization and download functions. The figure is created with BioRender.com.

Then, we classified cross-reactive epitopes based on experiment assays and receptor data. The identified epitope was considered cross-reactive if the sample used for the assay was exposed to immunogens (viral organism, protein or peptide) of a different origin or sequence than the candidate epitope. For epitopes known to bind receptors, those epitopes recognized by the same receptor were considered to be cross-reactive.

We further classified these epitopes based on viral origin and epitope type between the stimulatory immunogen and the identified positive epitope. Epitopes exhibiting cross-reactivity in different viral species were categorized as 'cross-viral' epitopes. Among them, cross-reactive immunogens with unknown sequences were termed undetermined, those with similar sequences and located in evolutionarily conserved regions were termed homologous and those with distinctly different sequences and positions were termed heterologous. Additionally, epitopes exhibiting cross-reactivity within the same virus species and its variants were categorized as 'cross-variant' epitopes. For cross-reactive immunogens, peptides with similar sequences and positions were termed homologous, while those with different sequences and positions were termed heterologous.

Antibodies collection and structure analysis

We retrieved the antibody list from the CoV-AbDab [38] database and searched PubMed using keywords such as 'identify SARS-CoV-2 antibodies', 'identify SARS-CoV antibodies', 'identify MERS-CoV antibodies' and 'identify coronavirus antibodies'. We performed data merging and preprocessing, resulting in a total of 12 613 binding and neutralizing antibodies derived from 571 references. For those antibodies with aliases, we unified and assigned a unique name for each antibody in the database and labeled aliases, facilitating searches using any antibody names.

Then, we manually collated 62 916 experimental data points on the affinity and neutralizing activity of these antibodies against various virus strains from the literature. These are determined by experimental measurements, with binding affinity assessed by surface plasmon resonance (SPR), biolayer interferometry (BLI) and enzyme-linked immunosorbent assay (ELISA) and neutralizing activity assessed by pseudovirus and live virus neutralization assays. The experimental measurements varied across studies, and to ensure objectivity, all experimental data were uniformly preprocessed and provided in the database. For binding affinity, the unit of the KD was unified as nanomoles (nM). KD less than 10 nM indicates high binding affinity, 10–1000 indicates moderate binding affinity and KD greater than 1000 indicates low binding affinity or even no binding. For neutralizing activity, the unit of half-maximum inhibitory concentration (IC_{50}) was unified as μg/ml. IC_{50} less than 1 μg/ml indicates strong neutralization, 1–10 indicates weak neutralization and greater than 10 indicates no neutralization.

For antibodies with resolved structures, we obtained the structures from the Protein Data Bank (https://www.rcsb.org/) [48], 988 in total. To study antigen–antibody binding and the effect of mutations on antibody binding, we utilized the 'InterfaceResidue' script in PyMOL [49] to calculate the binding sites of these antigen–antibody structural complexes. Residues on the two chains with a distance of less than 1 Å are considered interface contact residues. Then, at the site level, we depicted the functional annotations of antibody binding sites, indicating whether the site functions as an ACE-2 receptor binding site, has undergone mutations, is an mAb resistance mutation site and its conservation score across HCoVs.

Virological data collection and analysis for SARS-CoV-2 variants

To study the adaptive changes and immune evasion of SARS-CoV-2 variants during ongoing evolution, we organized current and past circulating variants according to Pango lineage [29]. Mutation sites in different sequences of the same variant might be different. The spike mutations of variants were obtained from the reputable database RCoV19 [27]. Then, we retrieved their virological characteristics on PubMed using keyword combinations such as 'SARS-CoV-2 virological characteristics', 'SARS-CoV-2 reproduction number', 'SARS-CoV-2 hACE2 bind', 'SARS-CoV-2 fusion assay' and 'SARS-CoV-2 infection assay'. This yielded 90 records of variant virological characteristics from 31 references, including transmissibility evaluated by basic reproduction number (R_0) and effective reproduction number (R_e), ACE2 binding affinity assessed by SPR and yeast surface display, fusogenicity assessed by cell-based fusion assay and infectivity assessed by cell-based virus infection assay. The experimental measurements varied across studies, and we preprocessed all experimental data uniformly and provided them in the database.

In addition, we collected 553 variant structures by keyword search in the Protein Data Bank (https://www.rcsb.org/) [48] and used Mol* Viewer [50] and iCn3D [51] for three-dimensional (3D) view. We further collected *in vitro* activity of these variants against advanced therapeutics from the National Center for Advancing Translational Sciences (NCATS) [52] with 10 158 data records covering 55 vaccines, 34 antibodies and 22 convalescent plasma samples. Each record includes drug class, drug name, viral mutation site, activity fold change and source literature. The *in vitro* activity was represented by IC_{50}, with smaller IC_{50} values indicating higher activity and larger IC_{50} values indicating lower activity. We took the negative logarithm (base 10) of the ratio of IC_{50} values compared to the wild-type (WT) or previous variants. Positive values indicate enhanced activity, whereas negative values indicate diminished activity compared to the WT or previous variants.

At the site level, we depicted the functional annotation of spike mutations, indicating whether the site serves as an ACE2 receptor binding site, is an mAb resistance mutation site and its conservation score across HCoVs. We identified the region of receptor binding domain (RBD) at residues 331–524 of the S protein and the region that binds to the ACE2 receptor at residues 428–508. The mAb resistance mutations were defined as spike mutations that met one or more of the following criteria: (1) having a median $\geq$ 5-fold reduction in susceptibility to a clinical stage mAb compared with WT according to CoV-RDB [34] and/or (2) having >0.1 average antibody escape from aggregated deep mutational scanning (DMS) data [53]. Additionally, we used the DMS results [54] to measure how RBD amino acid mutations affect protein expression and ACE2-binding affinity. DMS represents expression changes as Δlog(MFI), the mean fluorescence intensity of each variant relative to the unmutated SARS-CoV-2 RBD, with a positive value indicating increased expression, and represents binding affinity as Δlog10(KD, app), with positive values indicating stronger binding.

B-cell and T-cell epitope prediction pipeline

A large proportion of epitopes have not been experimentally validated. To provide potential epitopes, we constructed an epitope prediction pipeline (COVEP) that integrates existing recognized epitope prediction tools and provides a user-friendly application and web interface. It includes seven tools for B-cell epitope prediction, three tools for class I T-cell epitope prediction and two tools for class II T-cell epitope prediction. The protein sequences of HCoVs were downloaded from NCBI [26].

For the B-cell epitopes, seven tools embedded in the IEDB [44] were used for prediction, including BepiPred-1.0 [55], BepiPred-2.0 [56], Chou and Fasman beta turn prediction [57], Emini surface accessibility scale [58], Karplus and Schulz flexibility scale [59], Kolaskar and Tongaonkar antigenicity scale [60] and Parker hydrophilicity prediction [61]. These tools are commonly used and recognized for B-cell epitope prediction. Default parameters are used for each tool, and the thresholds for each tool are listed in Table S1. Amino acid sites confirmed by at least five tools are defined as part of B-cell epitopes in this study.

For T-cell epitopes, five tools were employed to predict two distinct types of T-cell epitopes: HLA class I alleles were predicted using NetMHCpan4.0 [62], MHCflurry2.0 [63] and DeepHLApan [64], while HLA class II alleles were predicted using MixMHC2pred [65] and NetMHCIIpan [66]. These tools have been validated for their accuracy in the literature of each tool and are widely used by researchers. Thresholds for defining potential T-cell epitopes for each tool are provided in Table S1. All HLA alleles available for T-cell epitope prediction in the COVEP pipeline are listed in Table S2. Additionally, we selected highly frequent HLA alleles (frequencies >5%) from different populations sourced from the 1000 Genome Project [67] and Zhou *et al.* [68] to predict potential T-cell epitopes in HCoVs. The specific alleles selected and their frequencies are given in Table S3. Only HLA-peptide pairs meeting all the tool thresholds are considered potential T-cell epitopes in HCoVs.

Database implementation

The web interface was developed using Django (v3.2.16) and JQuery (v3.5.1). MySQL (v5.7.34) was used for backend data storage and organization. Bootstrap [69] and Selectize [70] were utilized for UI components, and Datatables [71] was employed for visualizing search result tables. Statistical charts were created using ECharts [72]. The 3D view of PDB structures was displayed using iCn3D [51]. The g3-lollipop [73] was used to generate an interactive lollipop diagram for the mutation sites.

RESULTS

Comprehensive annotations and analysis of HCoVs B-cell and T-cell epitopes

The development of effective vaccines relies on the precise identification and characterization of immune epitopes. We collected and annotated 66 210 experimentally validated B-cell and 3209 T-cell epitopes of HCoVs in our database. Statistical analysis revealed that SARS-CoV and MERS-CoV have the most B-cell epitopes (about 15 000), followed by HCoV-OC43 and SARS-CoV-2 (over 9000), with other HCoVs having around 5000 (Figure 2A). Compared with others, SARS-CoV-2 has significantly the highest number of T-cell epitopes, totaling 2802 (Figure 2B). These epitopes are mainly derived from ORF1ab and the S protein.

During the SARS-CoV-2 pandemic, researchers conducted intensive studies and found that many peptides from other coronaviruses can trigger immune responses in samples from COVID-19 patients. These peptides are termed cross-viral reactive epitopes by us. About half of SARS-CoV-2 T-cell epitopes showed cross-reactivity with similar peptides in SARS-CoV-2 WT or variants, suggesting that these differences are not crucial for epitope recognition. These peptides are termed cross-variant reactive epitopes by us. See Methods for specific definitions and classifications. These classes are described in the 'Subclass' field of the epitope. The cross-active epitopes can elicit a broad

Figure 2. Immune epitope and antibody data statistics. (**A**) Number of B-cell epitopes for HCoVs and the proportion of source proteins and subclasses in each coronavirus. (**B**) Number of T-cell epitopes for HCoVs and the proportion of source proteins and subclasses in each coronavirus. (**C**) Number of binding and neutralizing antibodies against seven HCoVs. (**D**) Number of binding and neutralizing antibodies against SARS-CoV-2 variants.

immune response against viral strains and are important for the development of broad-spectrum vaccines.

The database interface offers a user-friendly search function, allowing queries based on epitope source, type and more. Users can also input peptides to retrieve all similar epitopes by BLAST or enter protein sequences to screen epitope fragments. Click on the epitope ID to access detailed annotations, including sequence, object type, location, source molecule, molecule parent, source organism, species, UniProt ID [74] and taxonomy ID [75]. The SARS-CoV-2 S protein undergoes frequent mutations, potentially resulting in antigenic drift or transfer and evasion of immune responses. We examined how sites in the epitope were mutated and compared variations among these variants. Figure 3 illustrates a representative details page. We constructed a heatmap with the horizontal axis representing mutations, the vertical axis representing variants and the value and color intensity indicating mutation frequency (Figure 3A). Conserved sites during coronavirus evolution are ideal targets for the development of broad-spectrum anti-SARS-CoV-2 drugs. We calculated the conservation scores of sites and epitopes in the homologous proteins of HCoVs and visualized the sequence alignment. The functional site annotation is displayed in the table, including mutations, receptor binding sites, mAb resistance sites and conserved sites. Click on the '+' for more details (Figure 3B). The details page describes the receptors that bind to the epitope, including encoding genes and sequence (Figure 3C). Click on the receptor ID to access all epitopes and experiments related to the receptor. The page also includes detailed entries for experimental assays, describing the protocols for immunostimulant procedures and experimental methods (Figure 3D). We labeled these receptors and experimental assays corresponding to the 'Subclass' field of the epitope. All experimental assays are accompanied by source references for researchers to explore further.

Molecular characteristics and interaction profiles of HCoV antibodies

The prevalence of SARS-CoV-2 variants has significantly reduced the effectiveness of existing antibodies, underscoring the importance of elucidating their molecular interactions and developing broadly neutralizing antibodies. The CovEpiAb database contains

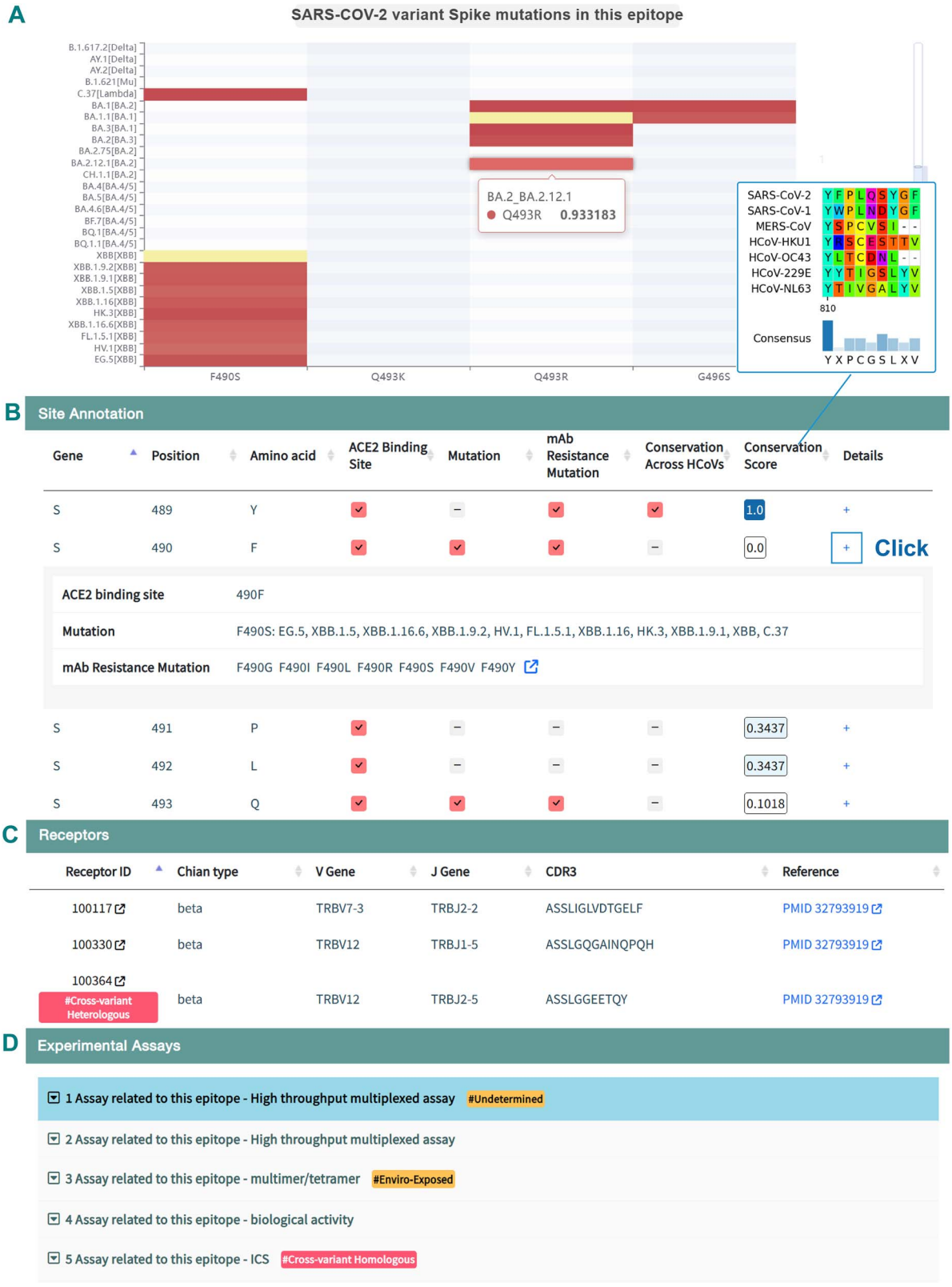

Figure 3. A typical epitope details page. (**A**) Spike mutations observed within epitope residues across various SARS-CoV-2 variants. (**B**) Alignment visualization and essential functional site annotations, encompassing receptor binding, mutations, mAb resistance and conservation. (**C**) Receptor data. (**D**) Experimental assays.

information on 12 613 antibodies against HCoVs and provides experimental interaction profiles related to binding and neutralization activity with 62 916 data points. Statistical analysis showed that 11 409 antibodies demonstrated binding to SARS-CoV-2 and 6710 showed neutralizing activity (Figure 2C). Over 3000 antibodies showed binding and neutralizing activity against SARS-CoV-2 variants, half of which were against Omicron variants (Figure 2D). In addition, we analyzed antibody gene usage and employed Sankey plots to illustrate the top 50 most frequent V and J gene pairs in both antibody heavy and light chains (Figure S2). The most commonly used human V and J genes for the heavy chain are IGKV3-30, IGHV1-69, IGHJ4 and IGHJ6 and for the light chain are IGKV1-39, IGKV3-20, IGKJ1 and IGLJ3.

Users can retrieve antibodies through three modules: antibody type, gene and structure, binding and neutralizing activity. Click on the antibody ID to access the antibody details page, which provides basic information including its name, type, source, binding viral antigen or region and structures, as well as binding and neutralizing characteristics against various viral strains. The structural analysis of antibody–antigen binding sites enables us to reveal molecular interactions and provide guidance for antibody design and optimization. Take the SA55 antibody as an example, which is in clinical trials and potently neutralizes various SARS-CoV-2 variants, including BQ.1.1 and XBB [76]. Figure 4A displays the resolved structures of SA55 with downloaded files and external links. Users can click the ID (green button) to view the 3D structure. Based on the complex structure, we identified the binding residues at the antigen–antibody contact interface. We focused on the antibody's target sites on the spike protein, as this is crucial to determine whether mutations in the spike protein impact antibody binding and function. We annotated mutations and key functional sites within the antibody binding region, including ACE2 binding sites, mAb resistance sites and conserved sites, with mouseover showing details.

In addition to molecular interactions, we collated experimental data on binding affinity and neutralizing activity against viral strains and mutation combinations to characterize the effectiveness of the antibody. The data are presented in a tabular format, making it easy to filter for records of interest, e.g. searching for BA.2 (Figure 4B). The results show that the neutralizing activity of the antibody was reduced against the BA.2 variant with an additional V503 mutation, while mutations at other sites have little effect. In conjunction with the structural analysis, 503 is a key binding site between spike protein and antibody heavy chain, and mutations at this site affect the antibody activity. The V503D mutation led to a complete loss of activity, which may be attributed to the effects of charge properties and spatial hindrance. The details page also includes information about the antibody's encoding genes, sequences and literature sources.

Virological characteristics of SARS-CoV-2 variants and therapeutic efficacy

SARS-CoV-2 genome mutations may lead to increased transmissibility and infectivity, evasion of immune response and reduced efficacy of vaccines and therapeutics. Therefore, intensive monitoring of the mutations and virological properties of emerging SARS-CoV-2 variants is required. Our database provides virological characteristics of SARS-CoV-2 variants, including current circulating Omicron subvariants and other variants such as B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma) and B.1.617.2 (Delta). To provide insight into the adaptive immune response against SARS-CoV-2, the 'Site Summary' page presents various resources together at the site level of SARS-CoV-2 spike protein, including antibody binding sites, experimentally validated and predicted T-cell and B-cell epitopes, spike mutations and conservation scores across HCoVs. Additionally, *in vitro* activity data of advanced therapeutics were collected, including 55 vaccines, 34 antibodies and 22 convalescent plasma samples, for a total of 10 158 data points. Statistical analysis revealed that the Omicron variants BQ and XBB exhibit a more significant reduction in neutralizing activity against various therapeutics compared with others. This reduction is particularly evident in neutralizing antibodies (Figure S3).

We constructed lollipop diagrams to illustrate and compare mutations on the spike protein of variants. Users can browse all variants on the 'Variants' page and click on 'Detail' to access the variant details page. Taking BA.4 as an example, Figure 5 is a typical details page. We annotated the key functional sites for spike mutations, including receptor binding sites, mAb resistance sites and conservation scores. For RBD, the effect of mutations on protein expression and receptor binding is also described based on DMS experimental results (Figure 5A). The emerging SARS-CoV-2 variants need to be carefully and rapidly assessed for their virological characteristics and drug resistance. The page provides information on changes in transmissibility, ACE2 binding affinity, fusogenicity and infectivity of the variant in comparison to others (Figure 5B). It also includes experimental methods, result descriptions and data sources. Additionally, we constructed a scatter plot to illustrate the variant's sensitivity to various therapeutics (Figure 5C). The *x*-axis represents therapeutics, and the *y*-axis represents the negative logarithm (base 10) of the fold change of *in vitro* IC_{50} values compared to the WT or previous variants. Positive values indicate enhanced neutralizing activity, while negative values indicate diminished activity. Toggle the buttons above the plot to explore different drug classes for effectiveness evaluation. Click on a point in the scatter plot to display detailed experimental information and references in the table below.

In the therapeutics panel of the 'Variant' page, the table displayed advanced vaccines and neutralizing antibodies, as well as convalescent plasma samples, including details about developers, approval status and clinical stage. Click on the 'Detail' button to enter the details page. In addition to the basic information, there is a similar scatter plot to quickly assess the neutralization activity of the therapeutic against various variants.

Epitope prediction pipeline and site conservation analysis

A large proportion of potential epitopes have not been validated, which emphasizes the importance of bioinformatics prediction in identifying and characterizing epitopes. Artificial intelligence (AI)–driven computational methods have been widely used for epitope prediction, especially for SARS-CoV-2 [77–79]. In this study, we developed COVEP, an integrated online pipeline for B-cell and T-cell epitope prediction (Figure 6A). The 'COVEP' page provides convenient access to the online web server, and the source code and docker image are available at GitHub and Docker Hub. Utilizing COVEP, we successfully predicted candidate B-cell and T-cell epitopes across the entire proteomes of seven HCoVs. For T-cell epitope prediction, we specifically selected highly frequent HLA alleles (frequencies >5%) from diverse populations (Table S3). The epitope prediction results can be found on the 'Predicted Epitopes' page. Table S4 shows the number of predicted epitopes in each protein of HCoVs.

To identify conserved sequences that may trigger a broad immune response, we calculated and analyzed the conserved region among five homologous proteins of SARS-CoV-2 and other

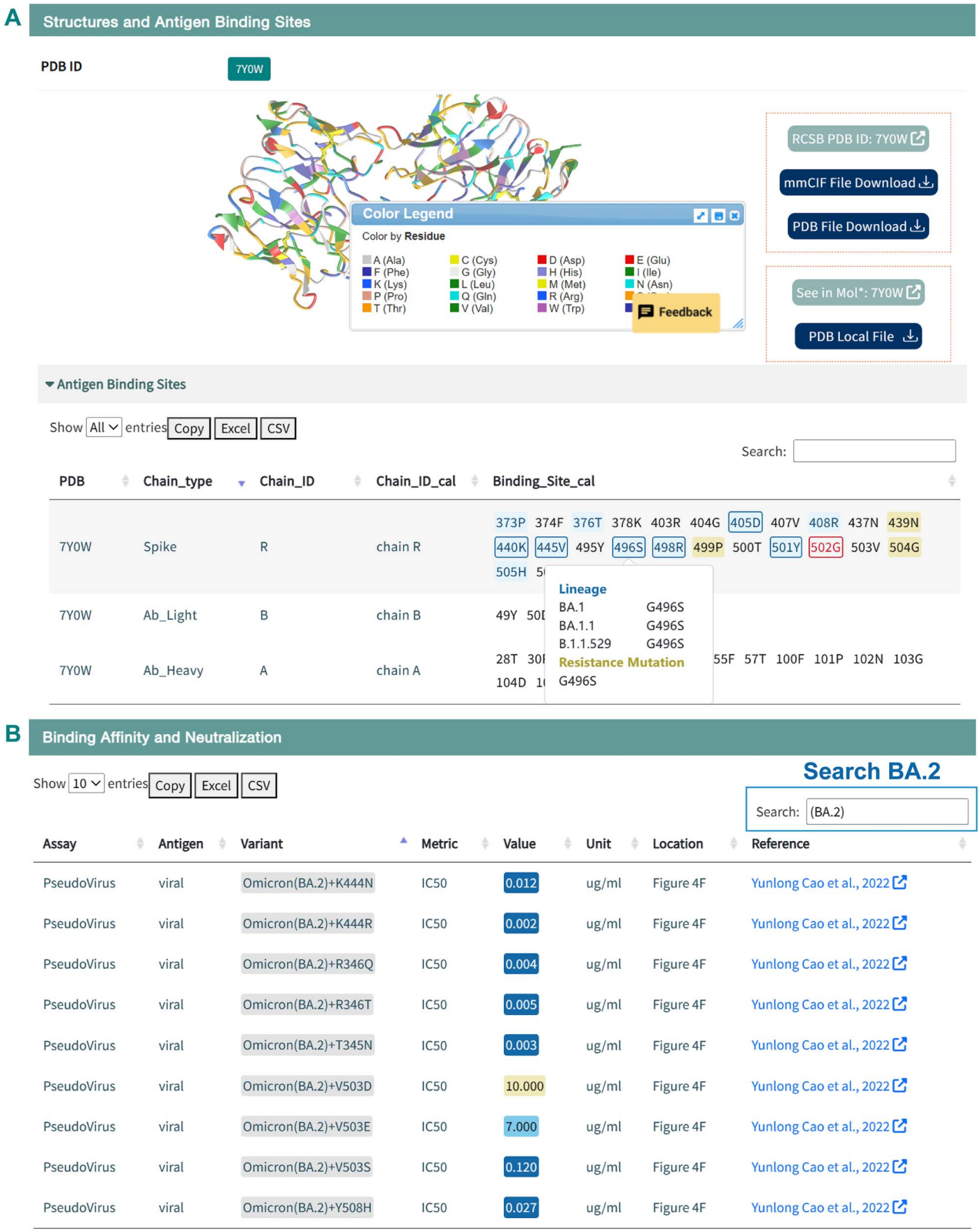

Figure 4. A typical antibody details page. (**A**) Resolved structures and binding sites. (**B**) Experimental data on binding affinity and neutralizing activity of the antibody.

HCoVs (Table S5). The most important is the S protein, and we identified nine conserved sequences on the S protein, all located on the S2 subunit (Figure 6B). They are clustered at sites 749–756, 815–826 (overlapping with the internal fusion peptide) and 915–1218 (including HR1). This is consistent with previously reported findings [80, 81], highlighting the conservation of the S2 subunit. Among them, the region 815–826 (RSFIEDLLFNKV) has been identified as the binding site for several broadly neutralizing

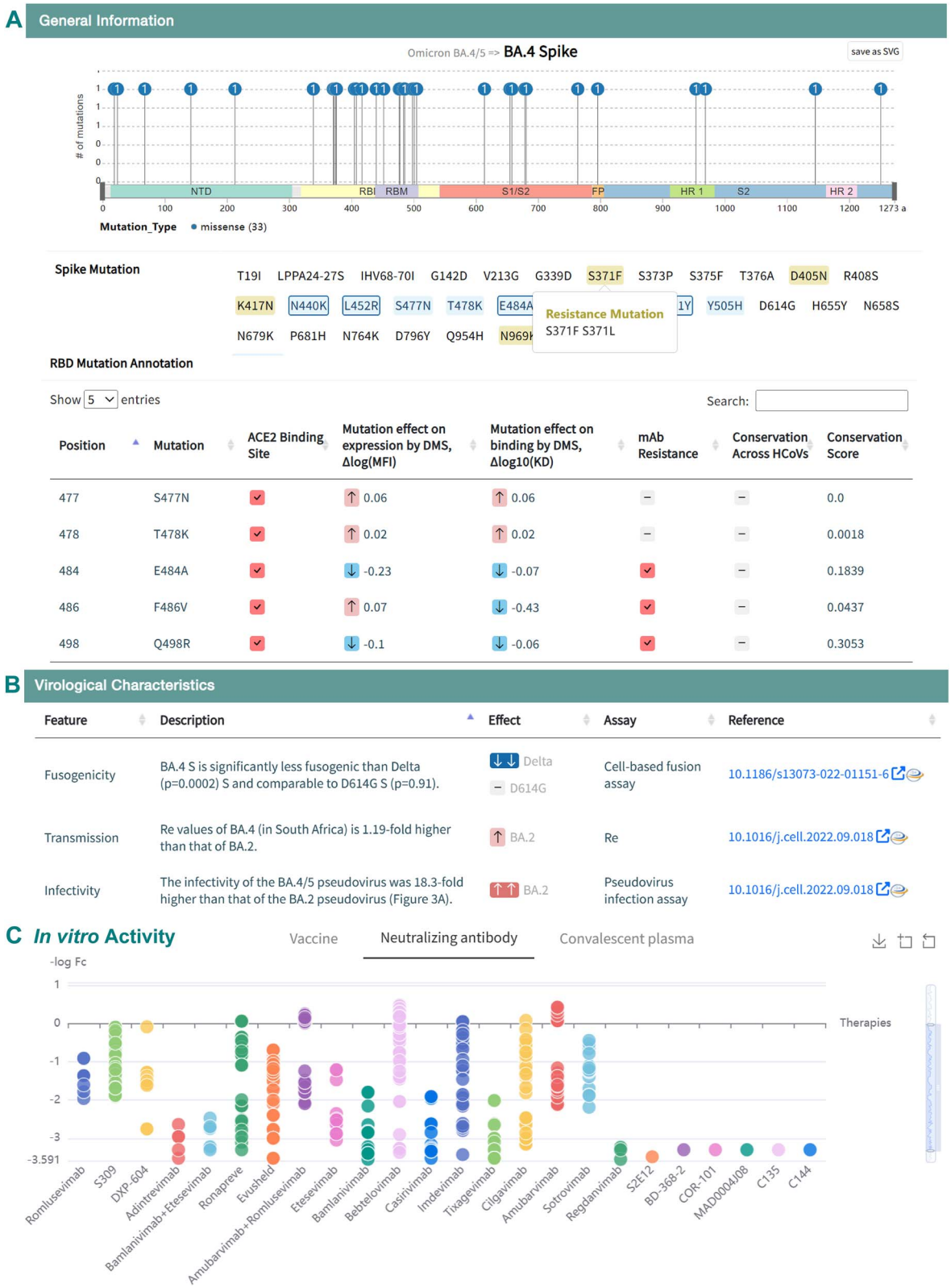

Figure 5. A typical SARS-CoV-2 variant details page. (**A**) Lollipop diagram and mutation annotation. (**B**) Virological characteristics including transmissibility, ACE2 binding affinity, fusogenicity and infectivity. (**C**) Scatter plot showing *in vitro* activity of various therapeutics against SARS-CoV-2 variants.

antibodies, such as COV44-62 and COV44-79. Key residues targeted by these broadly neutralizing antibodies are R815, E819, D820, L822 and F823 [24]. Additionally, we included information about antibodies on the 'HCoVs Conservation' page to visualize the spike protein conservation scores and antibody binding sites together. The epitopes within evolutionarily conserved regions represent crucial targets for the design of broad-spectrum vaccines against coronaviruses.

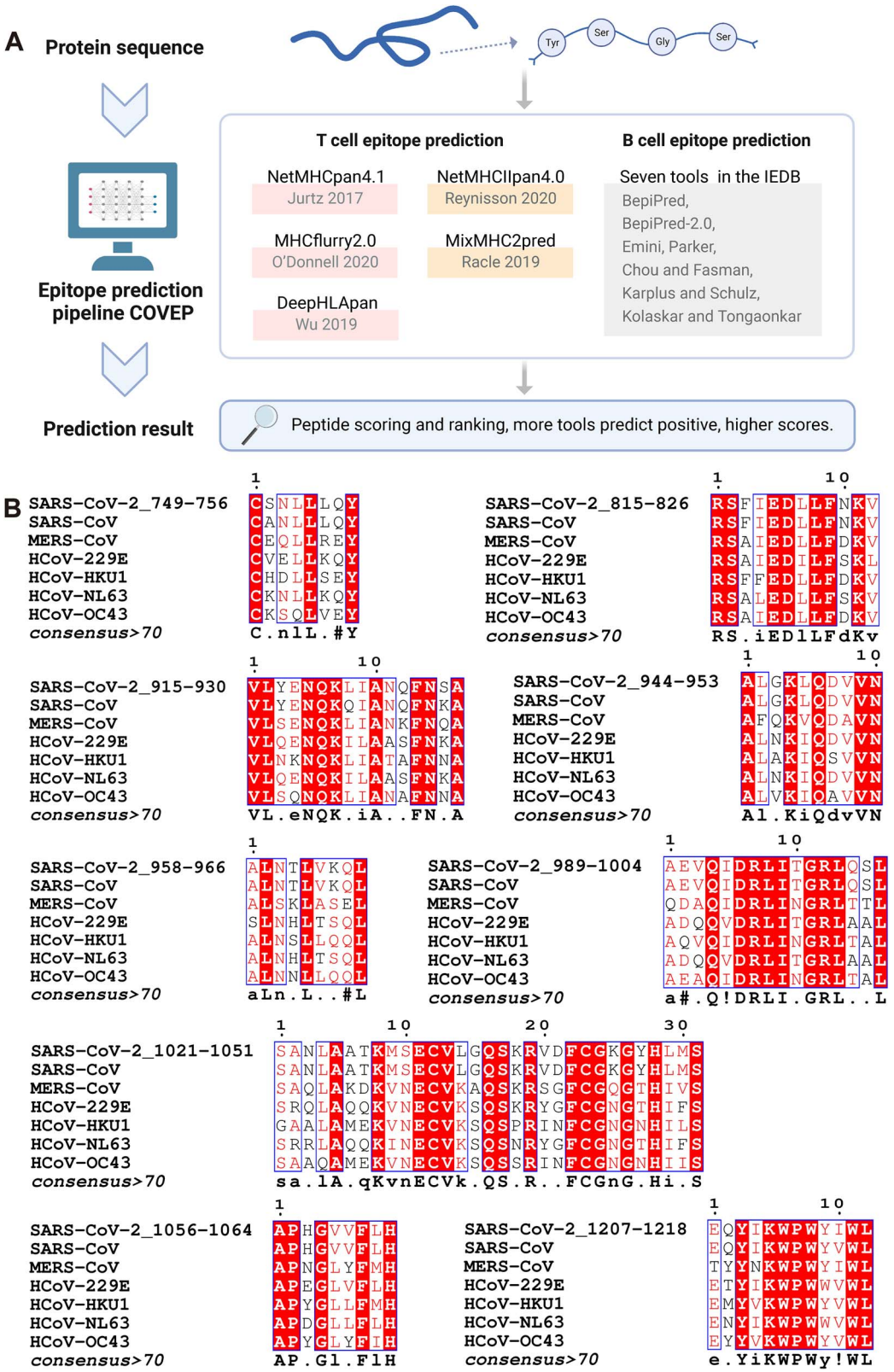

Figure 6. HCoVs epitope prediction and conservation analysis. (**A**) Workflow of the epitope prediction pipeline COVEP, including steps for B-cell and T-cell epitope prediction. (**B**) Multiple sequence alignments of conserved sequences in S protein of seven HCoVs.

CONCLUSIONS

In this study, we developed CovEpiAb, a functional database and analysis resource for HCoV immune epitopes and antibodies. The database contains an extensive collection, including information on over 60 000 epitopes and over 12 000 antibodies. Many peptides from different coronaviruses have been validated to induce immune responses in samples from COVID-19 patients. These epitopes have been systematically classified according to viral

origin and epitope type to describe cross-reactivity among different virus types and variants. For antibodies, we performed the structure-based analysis of antibody binding sites to elucidate molecular interactions. We collected approximately 70 000 data points to characterize the experimental interactions of antibodies on binding affinity and neutralizing activity. For SARS-CoV-2 variants, the database provides virological characteristics of the variants and *in vitro* activity data of various therapeutics. Additionally, at the site level, the database provides annotations of key functional sites for immune epitopes, antibody binding sites and mutations of SARS-CoV-2 variants. Another distinctive feature of CovEpiAb is the integrated online pipeline COVEP, which facilitates the identification of B-cell and T-cell epitopes. Our database is a unique and comprehensive resource for fitness and immune evasion assessment of emerging SARS-CoV-2 variants and for the design and development of broad-spectrum vaccines and neutralizing antibody drugs. We aim to update the database about every 3 months, and updates will continue indefinitely to reflect the ongoing progress of HCoV-related research.

Coronaviruses pose a substantial threat to human health, and although COVID-19 is no longer a major global concern, new coronaviruses may emerge and threaten humans. For viruses, it is essential to analyze the evolution of coronavirus and assess the immune escape caused by their mutations [82, 83]. Similarly, a thorough understanding of the dynamic immune responses and diverse outcomes caused by coronaviruses is necessary for individual immunity [84, 85]. Therefore, our future research will focus on analyzing antigenic epitope drift, immune cross-protection of emerging coronavirus variants and individual immune characteristics of virus-infected hosts. Combined with the analysis of coronavirus evolution and host immune interactions, this will help understand the dynamic immune response of the human body to viral infection and guide intervention therapeutics.

Key Points

- CovEpiAb is a database and analysis resource for HCoV epitopes and antibodies with an extensive collection of information on over 60 000 experimentally validated epitopes and 12 000 antibodies for HCoVs and SARS-CoV-2 variants.
- The database uniquely features cross-reactive epitope annotation, antibody molecular and experimental interaction profiles, virological characteristics and *in vitro* activity of SARS-CoV-2 variants.
- Site-level functional annotation of immune epitopes, antibodies and mutations illustrate how mutations affect epitope drift, antibody effectiveness and S protein function.
- The online integrated pipeline and conservation analysis allow users not only to predict candidate conserved epitopes for HCoVs but also for others, facilitating broad-spectrum vaccine design.

SUPPLEMENTARY DATA

Supplementary data are available online at http://bib.oxfordjournals.org/.

ACKNOWLEDGEMENTS

We thank the Information Technology Center and State Key Lab of CAD&CG, the Innovation Institute for Artificial Intelligence in Medicine, Zhejiang University and Alibaba Cloud for the support of computing resources. We also express our gratitude to all data contributors for providing the metadata that form the basis of this study. Additionally, we would like to acknowledge the group members and the numerous users who reported bugs and provided constructive suggestions.

AUTHOR CONTRIBUTIONS

Zhan Zhou (Conceptualization, funding acquisition, supervision, writing review and editing), Feng Zhu (Conceptualization, supervision, writing review and editing), Xue Zhang (data curation, formal analysis, investigation, methodology, visualization, writing original draft, writing review and editing), JingCheng Wu (data curation, formal analysis, investigation, methodology, visualization), Yuanyuan Luo (data curation, investigation), Yilin Wang (data curation, investigation), Yujie Wu (data curation, investigation), Xiaobin Xu (data curation, investigation), Yufang Zhang (data curation), Ruiying Kong (data curation), Ying Chi (data curation, methodology), Qiaojun He (funding acquisition, investigation), Yisheng Sun (investigation, methodology) and Shuqing Chen (investigation). All the authors have read and approved the final manuscript.

FUNDING

This work was supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No. LDT23H19011H19), the Huadong Medicine Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (Grant No. LHDMZ22H300002), the National Natural Science Foundation of China (Grant No. 31971371), K.P. Chao's High Technology Development Foundation of Zhejiang University (2022RC018) and the Scientific Research Fund of Zhejiang University (XY2022038).

DATA AVAILABILITY

All data in CovEpiAb are available at https://pgx.zju.edu.cn/covepiab. The COVEP webserver is available at https://pgx.zju.edu.cn/covepiab/tools/covep. The source code of COVEP is available on GitHub at https://github.com/zjupgx/COVEP, and the docker image is available on Docker Hub at https://hub.docker.com/r/biopharm/covep.

REFERENCES

1. Zhou P, Yang XL, Wang XG, *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;**579**(7798):270–3.
2. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int. [Accessed April 19, 2023].
3. Cao Y, Yisimayi A, Jian F, *et al.* BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by omicron infection. *Nature* 2022;**608**(7923):593–602.
4. Wang P, Nair MS, Liu L, *et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 2021;**593**(7857):130–5.
5. Wang Q, Iketani S, Li Z, *et al.* Alarming antibody evasion properties of rising SARS-CoV-2 BQ and XBB subvariants. *Cell* 2023;**186**(2):279–286.e8.

6. Miller J, Hachmann NP, Collier ARY, *et al.* Substantial neutralization escape by SARS-CoV-2 omicron variants BQ.1.1 and XBB.1. *N Engl J Med* 2023;**388**(7):662–4.
7. Tuekprakhon A, Nutalai R, Dijokaite-Guraliuc A, *et al.* Antibody escape of SARS-CoV-2 omicron BA.4 and BA.5 from vaccine and BA.1 serum. *Cell* 2022;**185**(14):2422–2433.e13.
8. Kurhade C, Zou J, Xia H, *et al.* Low neutralization of SARS-CoV-2 omicron BA.2.75.2, BQ.1.1 and XBB.1 by parental mRNA vaccine or a BA.5 bivalent booster. *Nat Med* 2023;**29**(2):344–7.
9. Garcia-Beltran WF, Lam EC, St Denis K, *et al.* Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* 2021;**184**(9):2372–2383.e9.
10. Anti-SARS-CoV-2 monoclonal antibodies. COVID-19 treatment guidelines. https://www.covid19treatmentguidelines.nih.gov/tables/variants-and-susceptibility-to-mabs/. [Accessed December 25, 2023].
11. Yamasoba D, Kosugi Y, Kimura I, *et al.* Neutralisation sensitivity of SARS-CoV-2 omicron subvariants to therapeutic monoclonal antibodies. *Lancet Infect Dis* 2022;**22**(7):942–3.
12. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 2016;**14**(8):523–34.
13. Stadler K, Masignani V, Eickmann M, *et al.* SARS—beginning to understand a new virus. *Nat Rev Microbiol* 2003;**1**(3):209–18.
14. Yan R, Zhang Y, Li Y, *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 2020;**367**(6485):1444–8.
15. Kesheh MM, Hosseini P, Soltani S, Zandi M. An overview on the seven pathogenic human coronaviruses. *Rev Med Virol* 2022;**32**(2):e2282.
16. Rucinski SL, Binnicker MJ, Thomas AS, Patel R. Seasonality of coronavirus 229E, HKU1, NL63, and OC43 from 2014 to 2020. *Mayo Clin Proc* 2020;**95**(8):1701–3.
17. Loyal L, Braun J, Henze L, *et al.* Cross-reactive CD4+ T cells enhance SARS-CoV-2 immune responses upon infection and vaccination. *Science* 2021;**374**(6564):eabh1823.
18. Kundu R, Narean JS, Wang L, *et al.* Cross-reactive memory T cells associate with protection against SARS-CoV-2 infection in COVID-19 contacts. *Nat Commun* 2022;**13**(1):80.
19. Le Bert N, Tan AT, Kunasegaran K, *et al.* SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* 2020;**584**(7821):457–62.
20. Shukla P, Pandey P, Prasad B, *et al.* Immuno-informatics analysis predicts B and T cell consensus epitopes for designing peptide vaccine against SARS-CoV-2 with 99.82% global population coverage. *Brief Bioinform* 2022;**23**(1):bbab496.
21. López D. Prediction of conserved HLA class I and class II epitopes from SARS-CoV-2 licensed vaccines supports T-cell cross-protection against SARS-CoV-1. *Biomedicine* 2022;**10**(7):1622.
22. Tajuelo A, López-Siles M, Más V, *et al.* Cross-recognition of SARS-CoV-2 B-cell epitopes with other Betacoronavirus nucleoproteins. *Int J Mol Sci* 2022;**23**(6):2977.
23. Tan CW, Chia WN, Young BE, *et al.* Pan-Sarbecovirus neutralizing antibodies in BNT162b2-immunized SARS-CoV-1 survivors. *N Engl J Med* 2021;**385**(15):1401–6.
24. Dacon C, Tucker C, Peng L, *et al.* Broadly neutralizing antibodies target the coronavirus fusion peptide. *Science* 2022;**377**(6607):728–35.
25. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data—from vision to reality. *Euro Surveill* 2017;**22**(13):30494.
26. Sayers EW, Bolton EE, Brister JR, *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2022;**50**(D1):D20–6.
27. Song S, Ma L, Zou D, *et al.* The global landscape of SARS-CoV-2 genomes, variants, and haplotypes in 2019nCoVR. *Genomics Proteomics Bioinformatics* 2020;**18**(6):749–59.
28. Hadfield J, Megill C, Bell SM, *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018;**34**(23):4121–3.
29. O'Toole Á, Hill V, Pybus OG, *et al.* Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch. *Wellcome Open Res* 2021;**6**:121.
30. Ferreira RC, Wong E, Gugan G, *et al.* CoVizu: rapid analysis and visualization of the global diversity of SARS-CoV-2 genomes. *Virus Evol* 2021;**7**(2):veab092.
31. Tsueng G, Mullen JL, Alkuzweny M, *et al.* Outbreak.info research library: a standardized, searchable platform to discover and explore COVID-19 resources. *Nat Methods* 2023;**20**(4):536–40.
32. Canakoglu A, Pinoli P, Bernasconi A, *et al.* ViruSurf: an integrated database to investigate viral sequences. *Nucleic Acids Res* 2021;**49**(D1):D817–24.
33. Fang S, Li K, Shen J, *et al.* GESS: a database of global evaluation of SARS-CoV-2/hCoV-19 sequences. *Nucleic Acids Res* 2021;**49**(D1):D706–14.
34. Tzou PL, Tao K, Pond SLK, Shafer RW. Coronavirus resistance database (CoV-RDB): SARS-CoV-2 susceptibility to monoclonal antibodies, convalescent plasma, and plasma from vaccinated persons. *PloS One* 2022;**17**(3):e0261045.
35. Sun Q, Shu C, Shi W, *et al.* VarEPS: an evaluation and prewarning system of known and virtual variations of SARS-CoV-2 genomes. *Nucleic Acids Res* 2022;**50**(D1):D888–97.
36. Gowthaman R, Guest JD, Yin R, *et al.* CoV3D: a database of high resolution coronavirus protein structures. *Nucleic Acids Res* 2021;**49**(D1):D282–7.
37. Wu J, Chen W, Zhou J, *et al.* COVIEdb: a database for potential immune epitopes of coronaviruses. *Front Pharmacol* 2020;**11**:572249.
38. Raybould MIJ, Kovaltsuk A, Marks C, Deane CM. CoV-AbDab: the coronavirus antibody database. *Bioinformatics* 2021;**37**(5):734–5.
39. Rawat P, Sharma D, Prabakaran R, *et al.* Ab-CoV: a curated database for binding affinity and neutralization profiles of coronavirus-related antibodies. *Bioinformatics* 2022;**38**(16):4051–2.
40. Yue Z, Zhang E, Xu C, *et al.* PAGER-CoV: a comprehensive collection of pathways, annotated gene-lists and gene signatures for coronavirus disease studies. *Nucleic Acids Res* 2021;**49**(D1):D589–99.
41. Amahong K, Zhang W, Zhou Y, *et al.* CovInter: interaction data between coronavirus RNAs and host proteins. *Nucleic Acids Res* 2023;**51**(D1):D546–56.
42. Chen TF, Chang YC, Hsiao Y, *et al.* DockCoV2: a drug database against SARS-CoV-2. *Nucleic Acids Res* 2021;**49**(D1):D1152–9.
43. Tworowski D, Gorohovski A, Mukherjee S, *et al.* COVID19 drug repository: text-mining the literature in search of putative COVID19 therapeutics. *Nucleic Acids Res* 2021;**49**(D1):D1113–21.
44. Vita R, Overton JA, Greenbaum JA, *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 2015;**43**(D1):D405–12.
45. Rozewicki J, Li S, Amada KM, *et al.* MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res* 2019;**47**(W1):W5–10.
46. Grant BJ, Rodrigues APC, ElSawy KM, *et al.* Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 2006;**22**(21):2695–6.

47. Shimoyama, Y. pyMSAviz: MSA visualization python package for sequence analysis [Computer software]. 2022. https://github.com/moshi4/pyMSAviz.
48. Berman HM, Westbrook J, Feng Z, *et al.* The protein data Bank. *Nucleic Acids Res* 2000;**28**(1):235–42.
49. PyMOL | pymol.org. https://pymol.org/2/. [Accessed July 11, 2023]
50. Sehnal D, Bittrich S, Deshpande M, *et al.* Mol∗ viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res* 2021;**49**(W1):W431–7.
51. Wang J, Youkharibache P, Zhang D, *et al.* iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics* 2020;**36**(1):131–5.
52. National Center for Advancing Translational Sciences. https://opendata.ncats.nih.gov/covid19/index.html. [Accessed February 28, 2023].
53. Starr TN, Greaney AJ, Addetia A, *et al.* Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* 2021;**371**(6531):850–4.
54. Starr TN, Greaney AJ, Hilton SK, *et al.* Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 2020;**182**(5):1295–1310.e20.
55. Larsen JEP, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res* 2006;**2**(1):2.
56. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017;**45**(W1):W24–9.
57. Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* 1978;**47**:45–148.
58. Emini EA, Hughes JV, Perlow DS, *et al.* Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 1985;**55**(3):836–9.
59. Karplus P, Schulz G. Prediction of chain flexibility in proteins—a tool for the selection of peptide antigens. *Naturwissenschaften* 1985;**72**(4):212–3.
60. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990;**276**(1–2):172–4.
61. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 1986;**25**(19):5425–32.
62. Jurtz V, Paul S, Andreatta M, *et al.* NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 2017;**199**(9):3360–8.
63. O'Donnell TJ, Rubinsteyn A, Laserson U. MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst* 2020;**11**(1):42–48.e7.
64. Wu J, Wang W, Zhang J, *et al.* DeepHLApan: a deep learning approach for Neoantigen prediction considering both HLA-peptide binding and immunogenicity. *Front Immunol* 2019;**10**:2559.
65. Racle J, Michaux J, Rockinger GA, *et al.* Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat Biotechnol* 2019;**37**(11):1283–6.
66. Reynisson B, Barra C, Kaabinejadian S, *et al.* Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *J Proteome Res* 2020;**19**(6):2304–15.
67. Gourraud PA, Khankhanian P, Cereb N, *et al.* HLA diversity in the 1000 genomes dataset. *PloS One* 2014;**9**(7):e97282.
68. Zhou F, Cao H, Zuo X, *et al.* Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nat Genet* 2016;**48**(7):740–6.
69. Bootstrap. https://getbootstrap.com/. [Accessed July 3, 2023].
70. Selectize. https://selectize.dev/. [Accessed July 3, 2023].
71. DataTables | Table plug-in for jQuery. https://datatables.net/. [Accessed July 3, 2023]
72. Li D, Mei H, Shen Y, *et al.* ECharts: a declarative framework for rapid construction of web-based visualization. *Vis Inform* 2018;**2**(2):136–46.
73. g3-lollipop.js. Interactively visualize genetic mutations using a lollipop-diagram. https://github.com/G3viz/g3lollipop.js.
74. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;**47**(D1):D506–15.
75. Federhen S. The NCBI taxonomy database. *Nucleic Acids Res* 2012;**40**(Database issue):D136–43.
76. Cao Y, Jian F, Zhang Z, *et al.* Rational identification of potent and broad sarbecovirus-neutralizing antibody cocktails from SARS convalescents. *Cell Rep* 2022;**41**(12):111845.
77. Grifoni A, Sidney J, Zhang Y, *et al.* A sequence homology and Bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* 2020;**27**(4):671–680.e2.
78. Mei X, Gu P, Shen C, *et al.* Computer-based Immunoinformatic analysis to predict candidate T-cell epitopes for SARS-CoV-2 vaccine design. *Front Immunol* 2022;**13**:847617.
79. Sohail MS, Ahmed SF, Quadeer AA, McKay MR. In silico T cell epitope identification for SARS-CoV-2: Progress and perspectives. *Adv Drug Deliv Rev* 2021;**171**:29–47.
80. Chai M, Guo Y, Yang L, *et al.* A high-throughput single cell-based antibody discovery approach against the full-length SARS-CoV-2 spike protein suggests a lack of neutralizing antibodies targeting the highly conserved S2 domain. *Brief Bioinform* 2022;**23**(3):bbac070.
81. Pang W, Lu Y, Zhao YB, *et al.* A variant-proof SARS-CoV-2 vaccine targeting HR1 domain in S2 subunit of spike protein. *Cell Res* 2022;**32**(12):1068–85.
82. Carabelli AM, Peacock TP, Thorne LG, *et al.* SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol* 2023;**21**(3):162–77.
83. Ma W, Fu H, Jian F, *et al.* Immune evasion and ACE2 binding affinity contribute to SARS-CoV-2 evolution. *Nat Ecol Evol* 2023;**7**(9):1457–66.
84. Feng Z, Ren X, Duren Z, Wang Y. Human genetic variants associated with COVID-19 severity are enriched in immune and epithelium regulatory networks. *Phenomics* 2022;**2**(6):389–403.
85. Nicholson JK. Molecular phenomic approaches to deconvolving the systemic effects of SARS-CoV-2 infection and post-acute COVID-19 syndrome. *Phenomics* 2021;**1**(4):143–50.