Computational and Structural Biotechnology Journal 21 (2023) 4697–4705

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Review article

Applying the digital data and the bioinformatics tools in SARS-CoV-2 research

Meng Tan [a,1], Jiaxin Xia [a,1], Haitao Luo [a,1], Geng Meng [b,*], Zhenglin Zhu [a,**]

[a] School of Life Sciences, Chongqing University, Chongqing, China
[b] College of Veterinary Medicine, China Agricultural University, Beijing, China

ARTICLE INFO

Keywords:
SARS-CoV-2
Bioinformatics tool
Database
Software
Webserver

ABSTRACT

Bioinformatics has been playing a crucial role in the scientific progress to fight against the pandemic of the coronavirus disease 2019 (COVID-19) caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The advances in novel algorithms, mega data technology, artificial intelligence and deep learning assisted the development of novel bioinformatics tools to analyze daily increasing SARS-CoV-2 data in the past years. These tools were applied in genomic analyses, evolutionary tracking, epidemiological analyses, protein structure interpretation, studies in virus-host interaction and clinical performance. To promote the *in-silico* analysis in the future, we conducted a review which summarized the databases, web services and software applied in SARS-CoV-2 research. Those digital resources applied in SARS-CoV-2 research may also potentially contribute to the research in other coronavirus and non-coronavirus viruses.

1. Introduction

To date, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1–4] has affected more than 8% population and caused over 6.8 million deaths worldwide (covid.observer, a website displaying the COVID-19 world statistics based on the data collected by the Johns Hopkins University Center for Systems Science and Engineering). With the increased importance of bioinformatics in biological and medical research [5–9], *in-silico* analysis has contributed greatly to SARS-CoV-2 research. As early as the first identification of SARS-CoV-2 in December 2019 [1], relevant bioinformatics analyses were performed thoroughly by multiple approaches [10–14]. The genomic and three-dimensional protein information of SARS-CoV-2 were clarified by sequencing and structural experiments [15,16]. SARS-CoV-2, belonging to *Coronaviridae*, is an enveloped plus-strand RNA virus with a genome encoding 16 non-structural, 4 structural and 9 accessory proteins [17,18]. Multiple sequencing technologies were applied extensively to determine the sequences of quickly evolving SARS-CoV-2 strains afterwards, resulting to an exponentially deposition of the SARS-CoV-2 genomic data. To properly store, assess and analyze the gigantic data, substantial bioinformatics tools were developed or updated. These relevant bioinformatics tools were applied in nearly all aspects of SARS-CoV-2 research, including sequence/protein information annotation, evolutionary/mutation analysis, epidemiological studies and therapy (drug/vaccine) development. With the purpose to provide a comprehensive perspective, we reviewed the recent progress in the bioinformatics tools applied to investigate SARS-CoV-2 and summarized the features and links of those referred databases/tools in figures (Figs. 1, 2) and tables (Tables S1-S4) to facilitate usage of these tools for researchers.

2. SARS-CoV-2 genetic sequence databases

For the pandemic of SARS-CoV-2 worldwide, the sequencing of SARS-CoV-2 genomes was performed globally [19,20]. The sequenced SARS-CoV-2 whole genomes are mostly stored and published in the Global Initiative on Sharing Avian Influenza Data (GISAID) [21], which serves as a rapid virus information sharing platform responsible for pandemic situations. As a traditional data-sharing platform, the NCBI nucleotide database only stores a small ratio of all published SARS-CoV-2 genomes [10]. However, the transcriptomic and epigenomic data associated with SARS-CoV-2/COVID-19 are mostly stored

* Correspondence to: College of Veterinary Medicine, China Agricultural University, Beijing 100094 China.
** Correspondence to: School of Life Sciences, Chongqing University, No.55 Daxuecheng South Road, Shapingba, Chongqing 401331, China.
E-mail addresses: mg@cau.edu.cn (G. Meng), zhuzl@cqu.edu.cn (Z. Zhu).
[1] These authors contributed equally.

https://doi.org/10.1016/j.csbj.2023.09.044
Received 12 July 2023; Received in revised form 29 September 2023; Accepted 29 September 2023
Available online 1 October 2023

in the Gene Expression Omnibus (GEO) dataset and the Sequence Read Archive (SRA) of NCBI. 2019nCoVR publishes the SARS-CoV-2 genomes sequenced by China institutes and provides a list referring the SARS-CoV-2 genomes protected by copyrights [22,23]. This database has been upgraded to version 4.0 and renamed RCoV19. In the updated version, this database also displays newly identified infections and mutations. The sequencing data of SARS-CoV-2 provided by Chinese institutes are mostly stored in the China National GeneBank DataBase (CNGBdb) [24]. For a better search and retrieving the virus sequence, the Virus Data Integration Platform (VirusDIP) compiles viral sequence data from NCBI, GISAID and CNGBdb. Those databases are updated frequently for the fast increase of SARS-CoV-2 data [25].

To facilitate the management of the SARS-CoV-2 data, traditional viral databases or online genome browsers incorporated the SARS-CoV-2 data into their analysis platforms. Several SARS-CoV-2 relevant integrative databases were developed. The NCBI Datasets Project built a user-friendly page (www.ncbi.nlm.nih.gov/sars-cov-2) to retrieve the SARS-CoV-2 data, including the genomes, proteins, CDS sequences, annotation and relevant reports. As of August 1, 2023, the SARS-CoV-2 resource page was redirected to the NCBI SARS-CoV-2 Virus Data Center. NIH provides comprehensive COVID-19 open-access data and computational resources in datascience.nih.gov/covid-19-open-access-resources. The Ensembl COVID-19 browser provides the annotation information of the SARS-CoV-2 genome [26]. As another general genomics

Fig. 1. The databases and tools applied in different aspects of SARS-CoV-2 research. The databases and the tools are differentiated by light purple rectangles and light green rectangles. The exemplary protein structure diagram is the SARS-CoV-2 spike glycoprotein (PDB ID, 6VXX).

database, the UCSC Genome Browser offers comprehensive sequence and annotation information and multi-use analysis tools relevant to SARS-CoV-2 [27–29]. A specialized page for SARS-CoV-2 was built in the traditional virology database, ViralZone [30], too. The Bacterial and Viral Bioinformatics Resource Center (BV-BRC) incorporated the SARS-CoV-2 into the online viral research platform [31]. In addition to providing sequence and literature resources, RCoV19, the UCSC Genome Browser, and BV-BRC also offer variant annotation and alignment features that can be used to track and identify emerging variants. The open-source project Nextstrain provides a personalized visualization of the phylogenetic trees and detailed annotation of variants of concerns (VOCs), important for comprehending the evolution and transmission of SARS-CoV-2 [32]. Nextstrain is popularized by its user-friendly data-access tools in the evolutionary analysis of SARS-CoV-2. COVID-19 Data Portal integrates resources on viral sequences, host sequences, expression, proteins, biochemistry, imaging and literature [33]. The Coronavirus Database (CoVdb) collected published coronavirus genomes and provides online tools for population genetics analysis and functional genomics analysis in a general coronavirus viewpoint [34].

The online tools provided by databases assisted the bioinformatics survey in SARS-CoV-2 research (Figs. 1, 2 and Supplemental Table S1). However, systematic or integrated *in-silico* SARS-CoV-2 research is mostly performed on local servers and relies on varieties of analysis software or bioinformatics tools (Supplemental Tables S2-S4). In terms of the research on SARS-CoV-2 genomics, HAVoC is a tool to assemble raw sequences and assign lineages [35]. The workflow ViralFlow is a recommended choice for Illumima pair-end sequencing data analysis and information processing [36]. The tool automates the reference-genome-based analysis pipeline, such as data processing, genome assembly, PANGO lineage assignment, mutation tracking, and intra-host variant screening. These features are also available in EDGE COVID-19 [37]. The resulted sequences and data can be sent to GenBank, GISAID, and INSDC. To generate high quality alignments, V-pipe provides a novel method, *ngshmmalign*, which is specialized for small and highly diversified genomes [38]. The tool supports quality control, read alignment, SNP identification and viral haplotype inference, too. COVID-Profiler provides both webserver and software to annotate

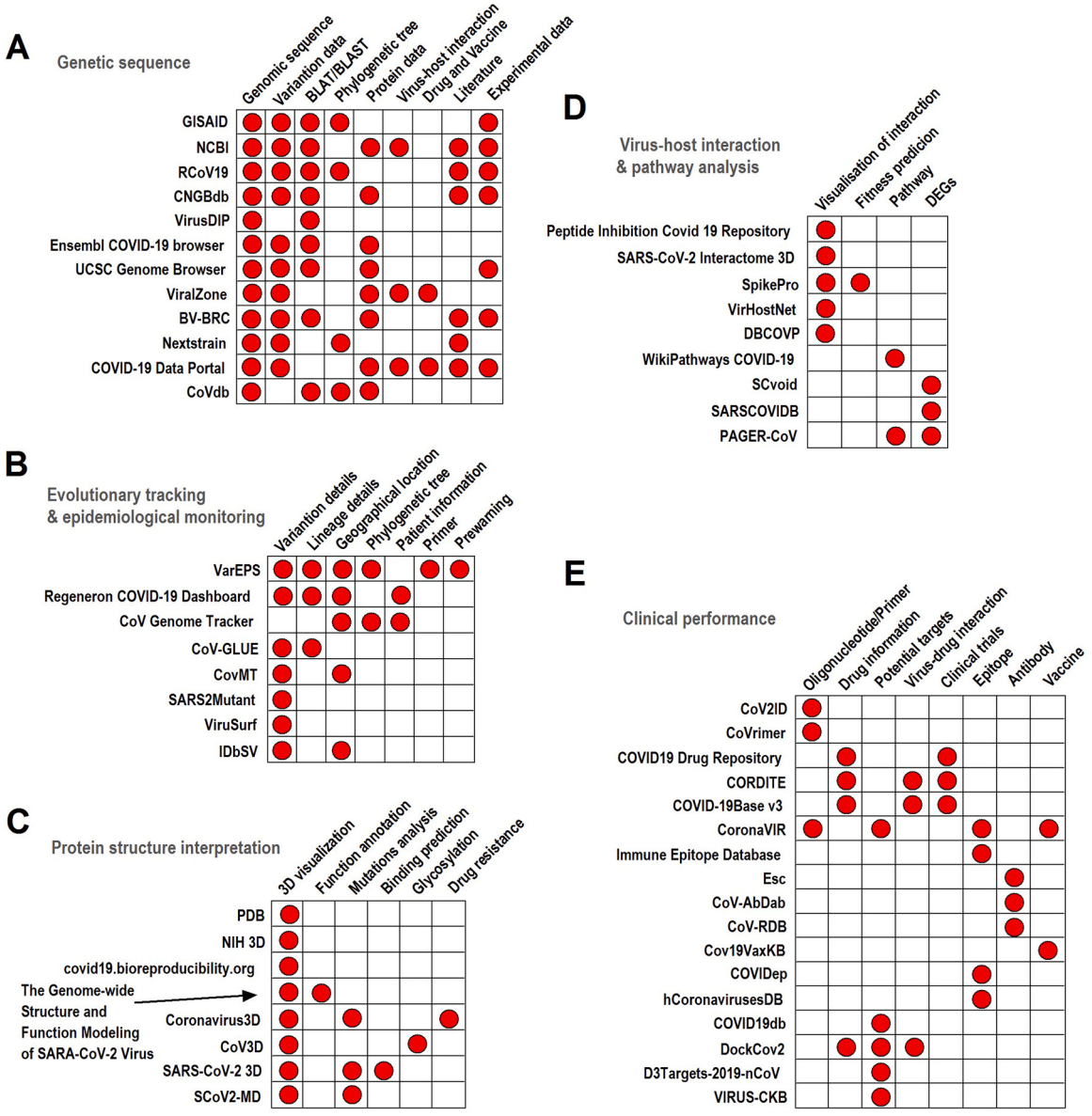

Fig. 2. The features provided by different databases. A red circle denotes that the feature (the corresponding item in the column) is provided by the database (the corresponding item in the row). 'DEGs' denotes differentially expressed genes in the subfigure D.

mutations from raw sequencing data and perform phylogenetic analysis [39]. The SARS-CoV-2 RECoVERY [40] and poreCov [41] are tools to construct viral genomes from raw sequencing data. The former additionally supports the analysis of variants. HAPHPIPE allows to perform a genome-wide assembly of viral consensus sequences and haplotypes, enabling rapid analysis of viral sequences generated by next-generation sequencing (NGS) platforms and providing high-quality output for downstream evolutionary analyses [42]. VADR provides a suite of tools to classify and analyze virus sequences, including norovirus, dengue fever, and SARS-CoV-2 viruses [43].

For the annotation of SARS-CoV-2 genome sequences, CoV-Seq is an online tool providing online tools for the analysis and visualization of the SARS-CoV-2 genome [44]. The tool automatically predicts genetic boundaries and identifies genetic variants. The tools, MicroGMT [45], CorGAT [46], BioAider [47], Coronapp [48], AutoVEM2 [49] and VirusViz [50], all allow to perform variant identification analysis. CorGAT focuses on variant function annotation, and AutoVEM2 additionally monitors haplotype subpopulations and prevalence trends. An online tool CoV-Spectrum is built by compiling and analyzing SARS-CoV-2 data from substantial sources, and allows to visualize information on variants and search for amino acid and nucleotide mutations [51]. COVID-Miner supports to identify information from genomic sequences and has a specific focus on the spike protein and the receptor binding domain [52]. The results are useful for vaccine design. Another online tool, the COVID-19 Viral Genome Analysis Pipeline, provides analysis tools that also focus on spike protein mutations [53]. For providing guidance of the vaccine design, CovDif is developed to detect the conserved region in the SARS-CoV-2 genome [54]. This tool also allows to identify conserved deletions due to point mutations. DeepIPs has a specialized deep learning architecture to find phosphorylation sites in host cells [55]. This tool assists the understanding of the molecular mechanisms of SARS-CoV-2 infection and change in host cell pathways.

3. Application in the evolutionary tracking and epidemiological monitoring of SARS-CoV-2 genetic variants

For epidemiological use, the technologies in big data, deep learning and artificial intelligence (AI) were applied to diagnose and monitor the epidemics of SARS-CoV-2 [56–58], enabling the reports of disease outbreaks in real-time [59] (Figs. 1, 2B and Table S3). Dedicated to facilitate the access and visualization of pandemic data, nCov2019 displays the data of SARS-CoV-2 strains along with the collection country and relevant clinical information [60]. As an early AI platform, BlueDot (bluedot.global) enables to detect the epidemic outbreak and visualize the spread of virus in real-time. The SARS-CoV-2 Variations Evaluation and Prewarning System (VarEPS) provides the prediction of the risk and spread of mutant strains using AI-based algorithm [61]. Based on the data from GISAID [21], the Regeneron COVID-19 Dashboard displays the geographical distribution of variants, details of mutations and corresponding information of patients. The CoV Genome Tracker is dedicated to tracking the Covid-19 epidemic with a haplotype network, a model that will be more accurate than phylogenetic tree [62]. For analyzing the effects of mutations, CoV-GLUE is a tool to annotate and analyze genome files with an emphasis on variations in amino acid sequences [63], while CovMT focuses on mutations in the RBD region of the virus [64]. SARS2Mutant allows to perform variant amino acid analysis based on numerous, high-quality SARS-CoV-2 protein sequences from GISAID [65]. ViruSurf [66] and International Database of SARS-CoV-2 Variations (IDbSV) [67] provide effective methods to search information regarding amino acid mutations and nucleotide variants.

The evolutionary analysis based on epidemiological and clinical data to identify the mutation pattern of SARS-CoV-2 is important for the prevention and control of SARS-CoV-2. In terms of building phylogenetic trees, PAUP [68] is frequently used [69–71] and the applied methods include the maximum parsimony and the maximum likelihood. FastTree is an effective tool in building large phylogenies and estimating the reliability [72,73]. Similar to FastTree, IQ-TREE is a maximum likelihood tool that enables a rapid assessment of various replacement models and choosing the best model for input sequences [74]. Beast performs well in the following evolutionary analysis steps [75]. This software applies the principles of Bayesian evolutionary analysis to estimate the phylogenetic relationship and the divergence time, and uses strict or relaxed molecular clock models to estimate rooted, time-measured phylogenies. For a simultaneous estimation of epidemiological parameters and pathogen phylogenies, the software PhyDyn, BEAST2 provides Bayesian phylogenetic inference using the models to deal with structured populations and complex population dynamics [76]. The Bayesian inference of phylogenies uses the Markov chain Monte Carlo (MCMC) and the results assist to understand the evolutionary history. The package Tracer is used to visualize and analyze MCMC trace files generated by Bayesian inference, including features such as kernel density estimation, multivariate visualization, and demographic trajectory reconstruction [77]. Another associated tool, Skygrowth, allows to use Bayesian and MCMC techniques to estimate phylodynamic inferences about effective population sizes for time-scale phylogenies [78]. The Susceptible-Exposed-Infectious-Removed model (SEIR) is wildly used in evolutionary analysis to predict the pandemic curve of SARS-CoV-2 [79].

The genetic subtypes of viruses are important for visualizing and analyzing geographical distribution. The Informative Subtype Markers (ISM) is a developed framework for the genetic subtyping of SARS-CoV-2 [80]. The result contains the regional differences in virus subtypes and the visualization of the subtypes that emerged at different times. The Genome Detective Coronavirus Typing Tool allows to identify the virus type, the genotype and the lineage of some nucleotide sequence [81]. Another tool, named Covidex, also allows to type the SARS-CoV-2 genome sequence [82]. Due to the large number of viral genomes, it is challenging to identify and assign lineages to gigantic SARS-CoV-2 sequences. For this issue, the software Pangolin is developed to assign the most likely lineage to a genomic sequence based on Pango nomenclature [83]. This software supports online and command line modes. Another online tool, CoVizu graphically displays the global genomic and lineage diversity of SARS-CoV-2, including the phylogenetic trees and evolutionary relationships between lineages [84]. Recombination may have a profound effect to the evolutionary process, and the detection of recombination events is essential to track and understand the evolutionary trajectory of viruses [85,86]. Concerning this issue, PoSeiDon is developed to provide an easy-to-use Nextflow pipeline to accurately detect positive selection and recombination events at specific points in the protein coding region [87]. Moreover, ViruClust is a tool to compare the spatiotemporal lineages and genome sequences of SARS-CoV-2, and supports the comparison of two sets of genomes and the prediction of lineage prevalence [88]. The results assist to identify possible variations. CovidPhy is a tool to process the sequencing data or accept identification codes stored in GISAID or GenBank, classify the genome into major phylogenetic nodes and provide information on the global frequency of viral variants and branches [89]. The effective phylogenetic tree optimization software matOptimize enables the optimization of the targeted SARS-CoV-2 phylogenies [90].

4. Protein structure interpretation

Except for the genomic and protein sequence level, studies on the protein structure of SARS-CoV-2 contribute to the development of structure-based therapies and vaccines. The non-structural, structural and accessory proteins participate in host cell entry, genome replication and transcription, and viral assembly and release [91]. Until now, the crystal structures of nearly all SARS-CoV-2 proteins have been resolved and published, according to the statistics provided by PDB COVID-19/SARS-CoV-2 Resources [92]. Analysis of the structures of

SARS-CoV-2 proteins helps to reveal the conformation, function and mechanism [93–96]. The molecular structure and function of SARS-CoV-2 may determine the possible antiviral medication. The Protein Data Bank (PDB) published the experimentally-determined 3D structure of the virus, including nucleic acids, proteins and polysaccharides [92]. So did NIH 3D, which developed a dedicated page for SARS-CoV-2 [97]. Based on those resources, the database Covid-19.bioreproducibility.org provide a validated and curated dataset of coronavirus protein structures, including 2942 SARS-CoV-2 protein structures and 206 protein structures of other coronaviruses [98]. The Genome-wide Structure and Function Modeling of SARS-CoV-2 Virus (for details, see Table S1) contains 3D structural models generated by the D-I-TASSER/ C-I-TASSER pipeline [99] and functional annotations of SARS-CoV-2 proteins. Coronavirus3D provides the integrated information of the protein 3D structures from PDB and the sequence mutation information from CNCB [100]. CoV3D provides a comprehensive dataset of SARS-CoV-2 protein structures and related complexes, and has a focus on the glycosylation of the spike protein [101]. The SARS-CoV-2 3D database combines the computationally predicted data and the experimentally validated data to enable the prediction of potential ligand binding and analysis of mutation sites [102]. The prediction method is based on the 3D structures and the prediction results is indicative for future relevant drug discovery. Moreover, understanding the relationship between 3D structure and dynamics is necessary to study how biological macromolecules work. SCoV2-MD produces simulations using a molecular dynamics method to investigate the structure-dynamics-function relationships of viral proteins [103]. Those structural databases/tools contribute to the structural analysis of SARS-CoV-2 proteins. Taking the investigation of a nonsynonymous mutation in some specific SARS-CoV-2 protein as an example, PDB, Covid-19.bioreproducibility.org or NIH 3D helps to retrieve the structural model of the protein. More information of the protein can be further retrieved from Coronavirus3D and CoV3D. Thereafter, we can predict the ligand binding and other effect influenced by the mutations through SARS-CoV-2 3D, as well as the dynamics of biological macromolecules using SCoV2-MD. Those initial analysis results may assist post analyses, such as protein-protein interaction through docking [104].

5. Virus-host interaction and pathway research

For the interaction between the RBD in the spike protein and the host receptor (ACE2) is critical in the invasion of virus, both the Peptide Inhibition Covid 19 Repository [105] and the SARS-CoV-2 Interactome 3D [106] visualize the interaction between RDB and ACE2, facilitating further exploration of this mechanism. Mutations present on the spike protein are likely relevant to viral adaptation and immune escape, SpikePro enables a rapid prediction of the adaptation of some mutant based on the structural models [107]. VirHostNet contains a relatively complete virus-host protein interaction resources that facilitates target determination and drug design [108]. Furthermore, DBCOVP is an important resource for experimental biologists engaged in coronavirus research studies and provides the complete repertoire, various sequence-structural properties and T-cell/B-cell epitopes of structural virulent glycoproteins from betacoronavirus [109].

As a traditional pathway database, the KEGG Pathway database displays the pathways associated with COVID-19 and coronavirus [110]. Another pathway database WikiPathways display the curated dataset concerning the molecular mechanisms of COVID-19 and the pathways related to SARS-CoV-2 [111]. Based on single-cell data, the database SCovid collected over 3000 significantly differentially expressed genes (DEGs) from ten human tissues [112]. SCovid can predict the essential genes and relevant potential therapeutics for different tissues. The prediction applies the machine learning technology. Another database, SARSCOVIDB, provide the analysis of the molecular impact of viral Infection by utilizing all infection-associated DEGs identified by literature-mining [113]. PAGER-CoV is a platform to search pathways associated with infection, inflammatory response and tissue repair [114]. The pathway search method is based on identified DEGs. The exhibited result is inductive for the identification of relevant molecular biological mechanisms and therapeutic approaches.

6. Clinical performance

In terms of the diagnosis, continuous SARS-CoV-2 mutation increases the difficulty to perform real-time virus detection, possibly resulting in ineffective treatment. For this issue, the database CoV2ID was developed to enable an effective analysis of oligonucleotide sequences [115]. The analysis method takes the genetic diversity of virus into consideration. Another resource, CoVrimer, collects published primer sequences and allow to change parameters to design potential primer pairs [116]. In terms of transmission monitoring, the tool A2B-COVID estimates the likelihood of infection between specific individuals and thus estimate and predict possible transmission events, through integrating the viral genomic data and information in the location of infected individuals [117]. CalmBelt is another tool that allows to identify patterns of outbreak transmission, detect potential variants, visualize the correlation of virulent strains, and track potential diagnostic escapes [118]. Transmission of the virus in different populations may vary in severity. Through automatic machine learning, an online platform COVIDOUTCOME is developed to estimate the disease severity based on sequence mutations and patient age [119]. In terms of clinical prediction, the Dynamic Nomogram is used to predict the prognosis of cancer patients with SARS-CoV-2 infection [120]. The information techniques have helped to shorten the time cost of diagnosis, improve the accuracy of diagnosis [121–125] and predict the transmission chain [126]. The Integrative CT Images and CFs for COVID-19 (iCTCF) is an open resource of chest computed tomography (CT) images and clinical features (CFs) for COVID-19, which is developed by Huazhong University of Science and Technology [127]. By using CT slices and CF data, iCTCF could accurately predict COVID-19 based on a convolutional neural network (CNN) model via deep learning.

The pandemic of SARS-CoV-2 used to make the development of drug and vaccine urgent. Molecular characterization and biological activity play a significant role in the study of drug discovery. Through assembling drug information from multiple literature resources, the COVID19 Drug Repository provides information on drug descriptions, side effects, publications, and pharmacological data [128]. Another database CORDITE provides detailed information on drug interactions, targets, clinical data, and publications [129]. A newly developed database, COVID-19Base, provides SARS-CoV-2 relevant information on disease, incorporating genes, miRNAs, drugs, side effects, and other factors [130]. iBioProVis allows to visualize the bioactive space of a compound and thus infer the potential target location of the compound of interest [131]. REDIAL-2020 is a package of tools to screen new compounds for anti-SRAS-CoV-2 activity by applying machine learning models [132]. Exploring the interaction between viruses and the human body also contributes the development of new drugs. Anh-Tien Ton et al. developed the Deep Docking (DD), a deep learning platform for structure-based virtual screening billions of molecular structures in a short time [133]. The COVID-19 Docking Server allows to predict the mode of binding of SARS-CoV-2 targets to ligands such as small molecules, peptides and antibodies, thus providing credibility for subsequent drug discovery [134]. In May 2023, the tool was updated to version 2.0 and renamed nCoVDock2, supporting more targets and adding a docking scoring feature [135]. The epitope prediction tool BepiPred-2.0 allows to predict the B-cell epitopes by selecting regions with high scores [136] and has been upgraded to version 3.0. Moreover, a series of bioinformatics tools were applied for antigenicity, physicochemical properties and protein structure prediction of the chimeric vaccine candidate. CRESSP uses the structural properties of proteins to identify cross-reactive epitopes in the SARS-CoV-2 and human proteomes [137]. This information serves as a foundation for further research into the

function of molecular mimicry in the post-infection diseases. Deep learning is supported by EMoMiS for evaluation of antigen-antibody binding, which provides helpful information for vaccine formulation and improvement [138]. CoVex collects information on virus-human protein interactions, human protein-protein interactions and drug-target interactions and is feasible to comprehend the molecular mechanisms of pathogenesis and evaluate potential therapy alternatives by investigating the virus-host-drug interactions [139].

An in-depth understanding of the mechanisms of immune escape is essential for the development of effective drugs and vaccines. A web-based resource CoronaVIR integrates several modules, including genomics, diagnosis, immunotherapy, drug designing, immunotherapy and drug design, and enables the screening of possible vaccine candidates and drug target information [140]. The Immune Epitope Database [141] allows to predict candidate targets for immune responses to SARS-CoV-2 [142]. Another database, ESC, reports SRAS-COV-2 variants related to potential antibody escape, in order to keep up with the rapid discovery of immune escape mechanisms [143]. Vaccine design requires a thorough understanding of the immunology systems and antibody data collection. The Oxford Protein Informatics Group developed CoV-AbDab, a collection of published or patented antibodies and nanobodies that bind to SARS-CoV-2 [144]. Monoclonal antibodies have both the capacity to divide continuously like the tumor cells and the capacity to produce antibodies like the immune cells. The Stanford Coronavirus Resistance Database (CoV-RDB) provides a neutralizing susceptibility data for SARS-CoV-2 mutations, SARS-CoV-2 monoclonal antibodies, recovery plasma and vaccine plasma variants [145]. This dataset helps to investigate the treatment of viral infection and the design of therapeutic regimens. Cov19VaxKB provides a web-based interface to search the information relevant to SARS-CoV-2/COVID-19 on vaccines, clinical trials, publications and vaccine adverse events [146]. This database also allows the statistical analysis and target prediction in vaccine design. COVIDep [147] and hCoronavirusesDB [148] enables the search for B cell and T cell epitopes for vaccine target development. COVIDep combines immunological data from SARS-CoV-2 and SRAS-CoV, while hCoronavirusesDB includes the sequence data as well as experimentally validated B cell and T cell epitope data.

For the prevention of SARS-CoV-2, molecular docking and associated algorithms are applied in drug discovery and target prediction. Through integrating substantial transcriptomic profiles relevant to SARS-COV-2, COVID19db provides a dataset of the drug-target-pathway interactions associated with COVID-19 [149]. The platform additionally offers analysis tools and drug development resources for locating prospective therapeutic targets at a transcriptomic level. DockCoV2 is a molecular docking-based resource and provides experimental data, pathway details, and enrichment analysis results [150]. This database also enables the prediction of the binding affinity of medications to proteins linked to stinger protein initiation, variant proteins, and human proteins. Another database, D3Targets-2019-nCoV, not only enables the prediction of drug targets from the drugs with experimental or clinical supports, but also enables virtual screening based on the structure of proteins to identify target compounds for potential drugs [151]. These should contribute to the advance in virology research and drug development. Based on well-established chemical genomics methods and analytical algorithms, Virus-CKB provides blood-brain barrier (BBB) prediction, docking to viral targets, and fingerprint-based similarity search [152]. These database utilities assist pharmacological research in potential therapeutic repurposing, drug combination, and drug-drug interaction (DDI) prediction.

7. Conclusion

Covering from genomic analysis to clinical performance, bioinformatics played an important role in SARS-CoV-2 research, although most part of the contribution is predictive and instructive. The validation of biological functions and clinical application predicted by bioinformatics approaches still need extensive experiments. Biological and medical experiments, as well as the sequencing, in turn would generate more data, which may induce new bioinformatics analyses and the creation of new bioinformatics tools based on the generated data. The digitalization of SARS-CoV-2 and the data mining based on the mega data, which has been performed outstandingly, is owing to the development of algorithm and informatics science in this digital age and this post-genomic era. The analysis of the gigantic biological data supports different aspects of SARS-CoV-2 research, and the establishment of a comprehensive data storage and analysis system is bound to the computational advances in the past three years. Hopefully, the databases, web service and standalone tools summarized in this review (Figs. 1, 2 and Tables S1-S4) will contribute to future studies of SARS-CoV-2 in theoretical aspects and clinical uses, moreover to research of general virology. Finally, we hope that this review would make it easier for researchers with different backgrounds to locate bioinformatics tools that are appropriate for the purpose of their research.

CRediT authorship contribution statement

Meng Tan, Jiaxin Xia, Haitao Luo, Geng Meng and Zhenglin Zhu took part in the writing of the manuscript. Zhenglin Zhu and Geng Meng conceived the idea and coordinated the project.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgments

We gratefully acknowledge the submitting and the originating laboratories where genetic sequence data were generated and shared via NCBI and the GISAID Initiative. This work was supported by grants from the National Natural Science Foundation of China (32170661), the National Key Research and Development Program (2019YFC1604600), the Fundamental Research Funds for the Central Universities (106112016CDJXY290002) and the National Natural Science Foundation of Hebei Province (19226631D).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.09.044.

References

[1] Ralph R, Lew J, Zeng T, Francis M, Xue B, Roux M, et al. 2019-nCoV (Wuhan virus), a novel Coronavirus: human-to-human transmission, travel-related cases, and vaccine readiness. J Infect Dev Ctries 2020;14(1):3–17. https://doi.org/10.3855/jidc.12425.

[2] Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan China: the mystery and the miracle. J Med Virol 2020;92(4):401–2. https://doi.org/10.1002/jmv.25678.

[3] Hui DS, E IA, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 2020;91:264–6. https://doi.org/10.1016/j.ijid.2020.01.009.

[4] Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579(7798):265–9. https://doi.org/10.1038/s41586-020-2008-3.

[5] Foulkes AC, Watson DS, Griffiths CEM, Warren RB, Huber W, Barnes MR. Research techniques made simple: bioinformatics for genome-scale biology. J Invest Dermatol 2017;137(9):e163–8. https://doi.org/10.1016/j.jid.2017.07.095.

[6] Fu Y, Ling Z, Arabnia H, Deng Y. Current trend and development in bioinformatics research. BMC Bioinforma 2020;21(Suppl 9):538. https://doi.org/10.1186/s12859-020-03874-y.

[7] Guo Y, Shen L, Shi X, Wang K, Dai Y, Zhao Z. Accelerating bioinformatics research with International Conference on Intelligent Biology and Medicine 2020. BMC Bioinforma 2020;21(Suppl 21):563. https://doi.org/10.1186/s12859-020-03890-y.

[8] Mulder NJ, Adebiyi E, Adebiyi M, Adeyemi S, Ahmed A, Ahmed R, et al. Development of Bioinformatics Infrastructure for Genomics Research. Glob Heart 2017;12(2):91–8. https://doi.org/10.1016/j.gheart.2017.01.005.
[9] van Kampen AH, Moerland PD. Taking bioinformatics to systems medicine. Methods Mol Biol 2016;1386:17–41. https://doi.org/10.1007/978-1-4939-3283-2_2.
[10] Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2020;48(D1):D9–16. https://doi.org/10.1093/nar/gkz899.
[11] Xiao K., Zhai J., Feng Y., Zhou N., Zhang X., Zou J.-J., et al. Isolation and Characterization of 2019-nCoV-like Coronavirus from Malayan Pangolins. bioRxiv, 2020; 2020:951335. https://doi.org/10.1101/2020.02.17.951335.
[12] Li X, Zai J, Zhao Q, Nie Q, Li Y, Foley BT, et al. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J Med Virol 2020;92(6):602–11. https://doi.org/10.1002/jmv.25731.
[13] Liu P, Chen W, Chen JP. Viral metagenomics revealed sendai virus and coronavirus infection of malayan pangolins (Manis javanica). Viruses 2019;11 (11). https://doi.org/10.3390/v11110979.
[14] Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev 2020;7(6):1012–23. https://doi.org/10.1093/nsr/nwaa036.
[15] Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 2020;9(1): 221–36. https://doi.org/10.1080/22221751.2020.1719902.
[16] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020;395(10224):565–74. https://doi.org/10.1016/S0140-6736(20)30251-8.
[17] Bai C, Zhong Q, Gao GF. Overview of SARS-CoV-2 genome-encoded proteins. Sci China-Life Sci 2022;65(2):280–94. https://doi.org/10.1007/s11427-021-1964-4.
[18] Arya R, Kumari S, Pandey B, Mistry H, Bihani SC, Das A, et al. Structural insights into SARS-CoV-2 proteins. J Mol Biol 2021;433(2):166725. https://doi.org/10.1016/j.jmb.2020.11.024.
[19] Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol 2020;92(6):667–74. https://doi.org/10.1002/jmv.25762.
[20] Phan T. Genetic diversity and evolution of SARS-CoV-2. Infect Genet Evol 2020; 81:104260. https://doi.org/10.1016/j.meegid.2020.104260.
[21] Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Eur Surveill 2017;22(13). https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.
[22] Bouhaddou M, Memon D, Meyer B, White KM, Rezelj VV, Correa Marrero M, et al. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. Cell 2020;182 (3):685–712. https://doi.org/10.1016/j.cell.2020.06.034.
[23] Zhao WM, Song SH, Chen ML, Zou D, Ma LN, Ma YK, et al. The 2019 novel coronavirus resource. Yi Chuan 2020;42(2):212–21. https://doi.org/10.16288/j.yczz.20-030.
[24] Chen FZ, You LJ, Yang F, Wang LN, Guo XQ, Gao F, et al. CNGBdb: China National GeneBank DataBase. Yi Chuan 2020;42(8):799–809. https://doi.org/10.16288/j.yczz.20-080.
[25] Wang L., Chen F., Guo X., You L.J., Yang X-x, Yang F., et al. VirusDIP: Virus Data Integration Platform. bioRxiv 2020;2020:139451. https://doi.org/10.1101/2020.06.08.139451.
[26] De Silva NH, Bhai J, Chakiachvili M, Contreras-Moreira B, Cummins C, Frankish A, et al. The Ensembl COVID-19 resource: ongoing integration of public SARS-CoV-2 data. Nucleic Acids Res 2022;50(D1):D765–70. https://doi.org/10.1093/nar/gkab889.
[27] Fernandes JD, Hinrichs AS, Clawson H, Gonzalez JN, Lee BT, Nassar LR, et al. The UCSC SARS-CoV-2 Genome Browser. Nat Genet 2020;52(10):991–8. https://doi.org/10.1038/s41588-020-0700-8.
[28] Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2021 update. Nucleic Acids Res 2021; 49(D1):D1046–57. https://doi.org/10.1093/nar/gkaa1070.
[29] Nassar LR, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2023 update. Nucleic Acids Res 2023;51(D1): D1188–95. https://doi.org/10.1093/nar/gkac1072.
[30] Masson P, Hulo C, De Castro E, Bitter H, Gruenbaum L, Essioux L, et al. ViralZone: recent updates to the virus knowledge resource. Nucleic Acids Res 2013;41(D1): D579–83. https://doi.org/10.1093/nar/gks1220.
[31] Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. Nucleic Acids Res 2023;51(D1):D678–89. https://doi.org/10.1093/nar/gkac1003.
[32] Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 2018;34(23): 4121–3. https://doi.org/10.1093/bioinformatics/bty407.
[33] Harrison PW, Lopez R, Rahman N, Allen SG, Aslam R, Buso N, et al. The COVID-19 Data Portal: accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing. Nucleic Acids Res 2021;49(W1):W619–23. https://doi.org/10.1093/nar/gkab417.
[34] Zhu Z, Meng K, Liu G, Meng G. A database resource and online analysis tools for coronaviruses on a historical and global scale. 2020:baaa070 Database (Oxf) 2020. https://doi.org/10.1093/database/baaa070.
[35] Truong Nguyen PT, Plyusnin I, Sironen T, Vapalahti O, Kant R, Smura T. HAVoC, a bioinformatic pipeline for reference-based consensus assembly and lineage assignment for SARS-CoV-2 sequences. BMC Bioinforma 2021;22(1):373. https://doi.org/10.1186/s12859-021-04294-2.
[36] Dezordi FZ, Neto A, Campos TL, Jeronimo PMC, Aksenen CF, Almeida SP, et al. ViralFlow: a versatile automated workflow for SARS-CoV-2 genome assembly, lineage assignment, mutations and intrahost variant detection. Viruses 2022;14 (2):217. https://doi.org/10.3390/v14020217.
[37] Lo CC, Shakya M, Connor R, Davenport K, Flynn M, Gutierrez AMY, et al. EDGE COVID-19: a web platform to generate submission-ready genomes from SARS-CoV-2 sequencing efforts. Bioinformatics 2022;38(10):2700–4. https://doi.org/10.1093/bioinformatics/btac176.
[38] Posada-Cespedes S, Seifert D, Topolsky I, Jablonski KP, Metzner KJ, Beerenwinkel N. V-pipe: a computational pipeline for assessing viral genetic diversity from high-throughput data. Bioinformatics 2021;37(12):1673–80. https://doi.org/10.1093/bioinformatics/btab015.
[39] Phelan J, Deelder W, Ward D, Campino S, Hibberd ML, Clark TG. COVID-profiler: a webserver for the analysis of SARS-CoV-2 sequencing data. BMC Bioinforma 2022;23(1):137. https://doi.org/10.1186/s12859-022-04632-y.
[40] De Sabato L., Vaccari G., Knijn A., Ianiro G., Di Bartolo I., Morabito S.Jb. SARS-CoV-2 RECoVERY: a multi-platform open-source bioinformatic pipeline for the automatic construction and analysis of SARS-CoV-2 genomes from NGS sequencing data. bioRxiv 2021;2021:425365. https://doi.org/10.1101/2021.01.16.425365.
[41] Brandt C, Krautwurst S, Spott R, Lohde M, Jundzill M, Marquet M, et al. poreCov-an easy to use, fast, and robust workflow for SARS-CoV-2 genome reconstruction via nanopore sequencing. Front Genet 2021;12:711437. https://doi.org/10.3389/fgene.2021.711437.
[42] Bendall ML, Gibson KM, Steiner MC, Rentia U, Perez-Losada M, Crandall KA. HAPHPIPE: haplotype reconstruction and phylodynamics for deep sequencing of intrahost viral populations. Mol Biol Evol 2021;38(4):1677–90. https://doi.org/10.1093/molbev/msaa315.
[43] Schäffer A.A., Hatcher E.L., Yankie L., Shonkwiler L., Brister J.R., Karsch-Mizrachi I., et al. VADR: validation and annotation of virus sequence submissions to GenBank. bioRxiv, 2020; 2020:852657. https://doi.org/10.1101/852657.
[44] Liu B., Liu K., Zhang H., Zhang L., Huang L. CoV-Seq: SARS-CoV-2 Genome Analysis and Visualization. bioRxiv, 2020;2020:071050. https://doi.org/10.1101/2020.05.01.071050.
[45] Xing Y, Li X, Gao X, Dong Q. MicroGMT: a mutation tracker for SARS-CoV-2 and other microbial genome sequences. Front Microbiol 2020;11:1502. https://doi.org/10.3389/fmicb.2020.01502.
[46] Chiara M, Zambelli F, Tangaro MA, Mandreoli P, Horner DS, Pesole G. CorGAT: a tool for the functional annotation of SARS-CoV-2 genomes. Bioinformatics 2021; 36(22–23):5522–3. https://doi.org/10.1093/bioinformatics/btaa1047.
[47] Zhu Z, Wang Y, Zhou X, Yang L, Meng G, Zhang Z. SWAV: a web-based visualization browser for sliding window analysis. Sci Rep 2020;10(1):149. https://doi.org/10.1038/s41598-019-57038-x.
[48] Mercatelli D, Triboli L, Fornasari E, Ray F, Giorgi FM. Coronapp: A web application to annotate and monitor SARS-CoV-2 mutations. J Med Virol 2021;93 (5):3238–45. https://doi.org/10.1002/jmv.26678.
[49] Xi B, Chen Z, Li S, Liu W, Jiang D, Bai Y, et al. AutoVEM2: A flexible automated tool to analyze candidate key mutations and epidemic trends for virus. Comput Struct Biotechnol J 2021;19:5029–38. https://doi.org/10.1016/j.csbj.2021.09.002.
[50] Bernasconi A, Gulino A, Alfonsi T, Canakoglu A, Pinoli P, Sandionigi A, et al. VirusViz: comparative analysis and effective visualization of viral nucleotide and amino acid variants. Nucleic Acids Res 2021;49(15):e90. https://doi.org/10.1093/nar/gkab478.
[51] Chen C, Nadeau S, Yared M, Voinov P, Xie N, Roemer C, et al. CoV-Spectrum: analysis of globally shared SARS-CoV-2 data to identify and characterize new variants. Bioinformatics 2022;38(6):1735–7. https://doi.org/10.1093/bioinformatics/btab856.
[52] Massacci A, Sperandio E, D'Ambrosio L, Maffei M, Palombo F, Aurisicchio L, et al. Design of a companion bioinformatic tool to detect the emergence and geographical distribution of SARS-CoV-2 Spike protein genetic variants. J Transl Med 2020;18(1):494. https://doi.org/10.1186/s12967-020-02675-4.
[53] Korber B, Fischer W, Gnanakaran S, Wyles M. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus (https://doi:) Cell 2020;182(4):812–27. https://doi.org/10.1016/j.cell.2020.06.043.
[54] Cedeno-Perez LF, Gomez-Romero L. CovDif, a Tool to Visualize the Conservation between SARS-CoV-2 Genomes and Variants. Viruses 2022;14(3):561. https://doi.org/10.3390/v14030561.
[55] Lv H, Dao FY, Zulfiqar H, Lin H. DeepIPs: comprehensive assessment and computational identification of phosphorylation sites of SARS-CoV-2 infection using a deep learning-based approach. Brief Bioinforma 2021;22(6):bbab244. https://doi.org/10.1093/bib/bbab244.
[56] Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How big data and artificial intelligence can help better manage the COVID-19 pandemic. Int J Environ Res Public Health 2020;17(9):3176. https://doi.org/10.3390/ijerph17093176.
[57] Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. Eur Respir J 2020;56: 2000775. https://doi.org/10.1183/13993003.00775-2020.
[58] Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes Metab Syndr: Clin Res Rev 2020;14(4):337–9. https://doi.org/10.1016/j.dsx.2020.04.012.

[59] Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020;368(6497):1362–7. https://doi.org/10.1126/science.abc0473.
[60] Wu T, Hu E, Ge X, Yu G. nCov2019: an R package for studying the COVID-19 coronavirus pandemic. PeerJ 2021;9:e11421. https://doi.org/10.7717/peerj.11421.
[61] Sun Q, Shu C, Shi W, Luo Y, Fan G, Nie J, 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. https://doi.org/10.1093/nar/gkab921.
[62] Akther S., Bezrucenkovas E., Sulkow B., Panlasigui C., Li L., Qiu W.-g, et al. CoV Genome Tracker: tracing genomic footprints of Covid-19 pandemic. bioRxiv, 2020; 036343. https://doi.org/10.1101/2020.04.10.036343.
[63] Singer JB, Gifford RJ, Cotten M, Robertson DL. CoV-GLUE: a web application for tracking SARS-CoV-2 genomic variation. 2020:2020060225 Preprints 2020;2020. https://doi.org/10.20944/preprints202006.0225.v1.
[64] Alam I, Radovanovic A, Incitti R, Kamau AA, Alarawi M, Azhar EI, et al. CovMT: an interactive SARS-CoV-2 mutation tracker, with a focus on critical variants. Lancet Infect Dis 2021;21(5):602. https://doi.org/10.1016/S1473-3099(21)00078-5.
[65] Rahimian K, Arefian E, Mahdavi B, Mahmanzar M, Kuehu DL, Deng Y. SARS2Mutant: SARS-CoV-2 amino-acid mutation atlas database. NAR Genom Bioinforma 2023;5(2):lqad037. https://doi.org/10.1093/nargab/lqad037.
[66] Canakoglu A, Pinoli P, Bernasconi A, Alfonsi T, Melidis DP, Ceri S. ViruSurf: an integrated database to investigate viral sequences. Nucleic Acids Res 2021;49(D1):D817–24. https://doi.org/10.1093/nar/gkaa846.
[67] Essabbar A, Kartti S, Alouane T, Hakmi M, Belyamani L, Ibrahimi A. IDbSV: An Open-Access Repository for Monitoring SARS-CoV-2 Variations and Evolution. Front Med 2021;8:765249. https://doi.org/10.3389/fmed.2021.765249.
[68] Swofford DL. PAUP*. phylogenetic analysis using parsimony (*and other methods), 4.0. beta 2002. https://doi.org/10.1002/0471650129.dob0522.
[69] Hu T, Li J, Zhou H, Li C, Holmes EC, Shi W. Bioinformatics resources for SARS-CoV-2 discovery and surveillance. Brief Bioinforma 2021;22(2):631–41. https://doi.org/10.1093/bib/bbaa386.
[70] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 2020;182(4):812–27. https://doi.org/10.1016/j.cell.2020.06.043.
[71] Li T, Liu D, Yang Y, Guo J, Feng Y, Zhang X, et al. Phylogenetic supertree reveals detailed evolution of SARS-CoV-2. Sci Rep 2020;10(1):22366. https://doi.org/10.1038/s41598-020-79484-8.
[72] Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;26(7):1641–50. https://doi.org/10.1093/molbev/msp077.
[73] Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 2010;5(3):e9490. https://doi.org/10.1371/journal.pone.0009490.
[74] Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 2015;32(1):268–74. https://doi.org/10.1093/molbev/msu300.
[75] Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolut Biol 2007;7:214. https://doi.org/10.1186/1471-2148-7-214.
[76] Volz EM, Siveroni I. Bayesian phylodynamic inference with complex models. PLoS Comput Biol 2018;14(11):e1006546. https://doi.org/10.1371/journal.pcbi.1006546.
[77] Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 2018;67(5):901–4. https://doi.org/10.1093/sysbio/syy032.
[78] Volz EM, Didelot X. Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. Syst Biol 2018;67(4):719–28. https://doi.org/10.1093/sysbio/syy007.
[79] Shurin MR, Morris A, Wells A, Wheeler SE. Assessing Immune Response to SARS-CoV-2 Infection. ImmunoTargets Ther 2020;9:111–4. https://doi.org/10.2147/ITT.S264138.
[80] Liu B, Liu K, Zhang H, Zhang L, Bian Y, Huang L. CoV-Seq, a New Tool for SARS-CoV-2 Genome Analysis and Visualization: Development and Usability Study. J Med Internet Res 2020;22(10):e22299. https://doi.org/10.2196/22299.
[81] Cleemput S., Dumon W., Fonseca V., Karim W.A., Giovanetti M., Alcantara L.C., et al. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. bioRxiv, 2020; 2020:928796. https://doi.org/10.1101/2020.01.31.928796.
[82] Cacciabue M, Aguilera P, Gismondi MI, Taboga O. Covidex: An ultrafast and accurate tool for SARS-CoV-2 subtyping. Infect Genet Evol 2022;99:105261. https://doi.org/10.1016/j.meegid.2022.105261.
[83] O'Toole A, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol 2021;7(2):veab064. https://doi.org/10.1093/ve/veab064.
[84] Ferreira RC, Wong E, Gugan G, Wade K, Liu M, Baena LM, et al. CoVizu: Rapid analysis and visualization of the global diversity of SARS-CoV-2 genomes. Virus Evol 2021;7(2):veab092. https://doi.org/10.1093/ve/veab092.
[85] Focosi D, Maggi F. Recombination in coronaviruses, with a focus on SARS-CoV-2. Viruses 2022;14(6):1239. https://doi.org/10.3390/v14061239.
[86] Zhu ZL, Meng KW, Meng G. Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. Sci Rep 2020;10:21617. https://doi.org/10.1038/S41598-020-78703-6.
[87] Holzer M, Marz M. PoSeiDon: a Nextflow pipeline for the detection of evolutionary recombination events and positive selection. Bioinformatics 2021;37(7):1018–20. https://doi.org/10.1093/bioinformatics/btaa695.
[88] Cilibrasi L, Pinoli P, Bernasconi A, Canakoglu A, Chiara M, Ceri S. ViruClust: direct comparison of SARS-CoV-2 genomes and genetic variants in space and time. Bioinformatics 2022;38(7):1988–94. https://doi.org/10.1093/bioinformatics/btac030.
[89] Bello X, Pardo-Seco J, Gomez-Carballa A, Weissensteiner H, Martinon-Torres F, Salas A. CovidPhy: A tool for phylogeographic analysis of SARS-CoV-2 variation. Environ Res 2022;204(A):111909. https://doi.org/10.1016/j.envres.2021.111909.
[90] Ye C, Thornlow B, Hinrichs A, Kramer A, Mirchandani C, Torvi D, et al. matOptimize: a parallel tree optimization method enables online phylogenetics for SARS-CoV-2. Bioinformatics 2022;38(15):3734–40. https://doi.org/10.1093/bioinformatics/btac401.
[91] Yang H, Rao Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. Nat Rev Microbiol 2021;19(11):685–700. https://doi.org/10.1038/s41579-021-00630-8.
[92] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235–42. https://doi.org/10.1093/nar/28.1.235.
[93] Turonova B, Sikora M, Schurmann C, Hagen WJH, Welsch S, Blanc FEC, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science 2020;370(6513):203–8. https://doi.org/10.1126/science.abd5223.
[94] Valcarcel A, Bensussen A, Alvarez-Buylla ER, Diaz J. Structural analysis of SARS-CoV-2 ORF8 protein: pathogenic and therapeutic implications. Front Genet 2021;12:693227. https://doi.org/10.3389/fgene.2021.693227.
[95] Abbasi BA, Saraf D, Sharma T, Sinha R, Singh S, Sood S, et al. Identification of vaccine targets & design of vaccine against SARS-CoV-2 coronavirus using computational and deep learning-based approaches. PeerJ 2022;10:e13380. https://doi.org/10.7717/peerj.13380.
[96] Zanchi FB, Mariuba LA, Nascimento V, Souza V, Corado A, Nascimento F, et al. Structural analysis of SARS-Cov-2 nonstructural protein 1 polymorphisms found in the Brazilian Amazon. Exp Biol Med 2021;246(21):2332–7. https://doi.org/10.1177/15353702211021348.
[97] Coakley MF, Hurt DE, Weber N, Mtingwa M, Fincher EC, Alekseyev V, et al. The NIH 3D print exchange: a public resource for bioscientific and biomedical 3d prints. 3D Print Addit Manuf 2014;1(3):137–40. https://doi.org/10.1089/3dp.2014.1503.
[98] Brzezinski D, Kowiel M, Cooper DR, Cymborowski M, Grabowski M, Wlodawer A, et al. Covid-19.bioreproducibility.org: A web resource for SARS-CoV-2-related structural models. Protein Sci 2021;30(1):115–24. https://doi.org/10.1002/pro.3959.
[99] Zheng W, Zhang C, Li Y, Pearce R, Bell EW, Zhang Y. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. Cell Rep Methods 2021;1(3):100014. https://doi.org/10.1016/j.crmeth.2021.100014.
[100] Sedova M, Jaroszewski L, Alisoltani A, Godzik A. Coronavirus3D: 3D structural visualization of COVID-19 genomic divergence. Bioinformatics 2020;36(15):4360–2. https://doi.org/10.1093/bioinformatics/btaa550.
[101] Gowthaman R, Guest JD, Yin R, Adolf-Bryfogle J, Schief WR, Pierce BG. CoV3D: a database of high resolution coronavirus protein structures. Nucleic Acids Res 2021;49(D1):D282–7. https://doi.org/10.1093/nar/gkaa731.
[102] Alsulami AF, Thomas SE, Jamasb AR, Beaudoin CA, Moghul I, Bannerman B, et al. SARS-CoV-2 3D database: understanding the coronavirus proteome and evaluating possible drug targets. Brief Bioinforma 2021;22(2):769–80. https://doi.org/10.1093/bib/bbaa404.
[103] Torrens-Fontanals M, Peralta-Garcia A, Talarico C, Guixa-Gonzalez R, Giorgino T, Selent J. SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions. Nucleic Acids Res 2022;50(D1):D858–66. https://doi.org/10.1093/nar/gkab977.
[104] Seeliger D, de Groot BL. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. J Comput-Aided Mol Des 2010;24(5):417–22. https://doi.org/10.1007/s10822-010-9352-6.
[105] Fernandez-Fuentes N, Molina R, Oliva B. A Collection of Designed Peptides to Target SARS-CoV-2 Spike RBD-ACE2 Interaction. Int J Mol Sci 2021;22(21):11627. https://doi.org/10.3390/ijms222111627.
[106] Ovek D, Taweel A, Abali Z, Tezsezen E, Koroglu YE, Tsai CJ, et al. SARS-CoV-2 Interactome 3D: A Web interface for 3D visualization and analysis of SARS-CoV-2-human mimicry and interactions. Bioinformatics 2022;38(5):1455–7. https://doi.org/10.1093/bioinformatics/btab799.
[107] Cia G, Kwasigroch JM, Rooman M, Pucci F. SpikePro: a webserver to predict the fitness of SARS-CoV-2 variants. Bioinformatics 2022;38(18):4418–9. https://doi.org/10.1093/bioinformatics/btac517.
[108] Guirimand T, Delmotte S, Navratil V. VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. Nucleic Acids Res 2015;43(D1):D583–7. https://doi.org/10.1093/nar/gku1121.
[109] Sahoo S, Mahapatra SR, Parida BK, Rath S, Dehury B, Raina V, et al. DBCOVP: A database of coronavirus virulent glycoproteins. Comput Biol Med 2021;129:104131. https://doi.org/10.1016/j.compbiomed.2020.104131.
[110] Arakawa K, Kono N, Yamada Y, Mori H, Tomita M. KEGG-based pathway visualization tool for complex omics data. Silico Biol 2005;5(4):419–23.
[111] Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways: connecting communities. Nucleic Acids Res 2021;49(D1):D613–21. https://doi.org/10.1093/nar/gkaa1024.

[112] Qi C, Wang C, Zhao L, Zhu Z, Wang P, Zhang S, et al. SCovid: single-cell atlases for exposing molecular characteristics of COVID-19 across 10 human tissues. Nucleic Acids Res 2022;50(D1):D867–74. https://doi.org/10.1093/nar/gkab881.
[113] da Rosa RL, Yang TS, Tureta EF, de Oliveira LRS, Moraes ANS, Tatara JM, et al. SARSCOVIDB-A New Platform for the Analysis of the Molecular Impact of SARS-CoV-2 Viral Infection. ACS Omega 2021;6(4):3238–43. https://doi.org/10.1021/acsomega.0c05701.
[114] Yue Z, Zhang E, Xu C, Khurana S, Batra N, Dang SDH, 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. https://doi.org/10.1093/nar/gkaa1094.
[115] Carneiro J., Gomes C., Couto C., Pereira F.Jb. CoV2ID: Detection and Therapeutics Oligo Database for SARS-CoV-2. 2020; 048991. bioRxiv https://doi.org/10.1101/2020.04.19.048991.
[116] Vural-Ozdeniz M, Akturk A, Demirdizen M, Leka R, Acar R, Konu O. CoVrimer: A tool for aligning SARS-CoV-2 primer sequences and selection of conserved/degenerate primers. Genomics 2021;113(5):3174–84. https://doi.org/10.1016/j.ygeno.2021.07.020.
[117] Illingworth CJR, Hamilton WL, Jackson C, Warne B, Popay A, Meredith L, et al. A2B-COVID: A Tool for Rapidly Evaluating potential SARS-CoV-2 transmission events. Mol Biol Evol 2022;39(3):msac025. https://doi.org/10.1093/molbev/msac025.
[118] Yingtaweesittikul H, Ko K, Abdul Rahman N, Tan SYL, Nagarajan N, Suphavilai C. CalmBelt: rapid SARS-CoV-2 genome characterization for outbreak tracking. Front Med 2021;8:790662. https://doi.org/10.3389/fmed.2021.790662.
[119] Nagy A, Ligeti B, Szebeni J, Pongor S, Gyrffy B. COVIDOUTCOME-estimating COVID severity based on mutation signatures in the SARS-CoV-2 genome. Database (Oxf) 2021;2021:baab020. https://doi.org/10.1093/database/baab020.
[120] Song C, Dong Z, Gong H, Liu XP, Dong X, Wang A, et al. An online tool for predicting the prognosis of cancer patients with SARS-CoV-2 infection: a multi-center study. J Cancer Res Clin Oncol 2021;147(4):1247–57. https://doi.org/10.1007/s00432-020-03420-6.
[121] Ko H, Chung H, Kang WS, Kim KW, Shin Y, Kang SJ, et al. COVID-19 Pneumonia Diagnosis Using a Simple 2D Deep Learning Framework With a Single Chest CT Image: Model Development and Validation. J Med Internet Res 2020;22(6):e19569. https://doi.org/10.2196/19569.
[122] Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. Cell 2020;182(5):1360. https://doi.org/10.1016/j.cell.2020.08.029.
[123] Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. Nat Commun 2020;11(1):3543. https://doi.org/10.1038/s41467-020-17280-8.
[124] Kuchana M, Srivastava A, Das R, Mathew J, Mishra A, Khatter K. AI aiding in diagnosing, tracking recovery of COVID-19 using deep learning on Chest CT scans. Multimed Tools Appl 2020;80:9161–75. https://doi.org/10.1007/s11042-020-10010-8.
[125] Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020;121:103792. https://doi.org/10.1016/j.compbiomed.2020.103792.
[126] Ye Q, Zhou J, Wu H. Using Information Technology to Manage the COVID-19 Pandemic: Development of a Technical Framework Based on Practical Experience in China. JMIR Med Inform 2020;8(6):e19515. https://doi.org/10.2196/19515.
[127] Ning W, Lei S, Yang J, Cao Y, Jiang P, Yang Q, et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. Nat Biomed Eng 2020;4(12):1197–207. https://doi.org/10.1038/s41551-020-00633-5.
[128] Tworowski D, Gorohovski A, Mukherjee S, Carmi G, Levy E, Detroja R, et al. COVID19 Drug Repository: text-mining the literature in search of putative COVID19 therapeutics. Nucleic Acids Res 2021;49(D1):D1113–21. https://doi.org/10.1093/nar/gkaa969.
[129] Martin R, Lochel HF, Welzel M, Hattab G, Hauschild AC, Heider D. CORDITE: The Curated CORona Drug InTERactions Database for SARS-CoV-2. iScience 2020;23(7):101297. https://doi.org/10.1016/j.isci.2020.101297.
[130] Basit SA, Qureshi R, Musleh S, Guler R, Rahman MS, Biswas KH, et al. COVID-19Base v3: Update of the knowledgebase for drugs and biomedical entities linked to COVID-19. Front Public Health 2023;11:1125917. https://doi.org/10.3389/fpubh.2023.1125917.
[131] Donmez A, Rifaioglu AS, Acar A, Dogan T, Cetin-Atalay R, Atalay V. iBioProVis: interactive visualization and analysis of compound bioactivity space. Bioinformatics 2020;36(14):4227–30. https://doi.org/10.1093/bioinformatics/btaa496.
[132] Kc GB, Bocci G, Verma S, Hassan MM, Holmes J, Yang JJ, et al. A machine learning platform to estimate anti-SARS-CoV-2 activities. Nat Mach Intell 2021;3(6):527–35. https://doi.org/10.1038/s42256-021-00335-w.
[133] Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3Billion Compounds. Mol Inform 2020;39(8):e2000028. https://doi.org/10.1002/minf.202000028.
[134] Kong R, Yang G, Xue R, Liu M, Wang F, Hu J, et al. COVID-19 Docking Server: a meta server for docking small molecules, peptides and antibodies against potential targets of COVID-19. Bioinformatics 2020;36(20):5109–11. https://doi.org/10.1093/bioinformatics/btaa645.
[135] Liu K, Lu X, Shi H, Xu X, Kong R, Chang S. nCoVDock2: a docking server to predict the binding modes between COVID-19 targets and its potential ligands. Nucleic Acids Res 2023;51(W1):W365–71. https://doi.org/10.1093/nar/gkad414.
[136] 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. https://doi.org/10.1093/nar/gkx346.
[137] An H, Eun M, Yi J, Park J. CRESSP: a comprehensive pipeline for prediction of immunopathogenic SARS-CoV-2 epitopes using structural properties of proteins. Brief Bioinforma 2022;23(2):bbac056. https://doi.org/10.1093/bib/bbac056.
[138] Stebliankin V., Baral P., Balbin C.A., Nunez-Castilla J., Sobhan M., Cickovski T. M., et al. EMoMiS: A Pipeline for Epitope-based Molecular Mimicry Search in Protein Structures with Applications to SARS-CoV-2. BioRxiv, 2022; 2022: 479274. https://doi.org/10.1101/2022.02.05.479274.
[139] Sadegh S, Matschinske J, Blumenthal DB, Galindez G, Kacprowski T, List M, et al. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. Nat Commun 2020;11(1):3518. https://doi.org/10.1038/s41467-020-17189-2.
[140] Patiyal S, Kaur D, Kaur H, Sharma N, Dhall A, Sahai S, et al. A web-based platform on coronavirus disease-19 to maintain predicted diagnostic, drug, and vaccine candidates. Monoclon Antibodies Immunodiagn Immunother 2020;39(6):204–16. https://doi.org/10.1089/mab.2020.0035.
[141] Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res 2015;43(D1):D405–12. https://doi.org/10.1093/nar/gku938.
[142] Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe 2020;27(4):671–80. https://doi.org/10.1016/j.chom.2020.03.002.
[143] Rophina M, Pandhare K, Shamnath A, Imran M, Jolly B, Scaria V. ESC: a comprehensive resource for SARS-CoV-2 immune escape variants. Nucleic Acids Res 2022;50(D1):D771–6. https://doi.org/10.1093/nar/gkab895.
[144] Raybould MIJ, Kovaltsuk A, Marks C, Deane CM. CoV-AbDab: the coronavirus antibody database. Bioinformatics 2021;37(5):734–5. https://doi.org/10.1093/bioinformatics/btaa739.
[145] 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. https://doi.org/10.1371/journal.pone.0261045.
[146] Huang PC, Goru R, Huffman A, Yu Lin A, Cooke MF, He Y. Cov19VaxKB: A Web-based Integrative COVID-19 Vaccine Knowledge Base. Vaccin: X 2021;10:100139. https://doi.org/10.1016/j.jvacx.2021.100139.
[147] Ahmed SF, Quadeer AA, McKay MR. COVIDep: a web-based platform for real-time reporting of vaccine target recommendations for SARS-CoV-2. Nat Protoc 2020;15(7):2141–2. https://doi.org/10.1038/s41596-020-0358-9.
[148] Almansour I, Boudellioua I. hCoronavirusesDB: an integrated bioinformatics resource for human coronaviruses. Database (Oxf) 2022;2022:baac017. https://doi.org/10.1093/database/baac017.
[149] Zhang W, Zhang Y, Min Z, Mo J, Ju Z, Guan W, et al. COVID19db: a comprehensive database platform to discover potential drugs and targets of COVID-19 at whole transcriptomic scale. Nucleic Acids Res 2022;50(D1):D747–57. https://doi.org/10.1093/nar/gkab850.
[150] Chen TF, Chang YC, Hsiao Y, Lee KH, Hsiao YC, Lin YH, et al. DockCoV2: a drug database against SARS-CoV-2. Nucleic Acids Res 2021;49(D1):D1152–9. https://doi.org/10.1093/nar/gkaa861.
[151] Shi Y, Zhang X, Mu K, Peng C, Zhu Z, Wang X, et al. D3Targets-2019-nCoV: a webserver for predicting drug targets and for multi-target and multi-site based virtual screening against COVID-19. Acta Pharm Sin B 2020;10(7):1239–48. https://doi.org/10.1016/j.apsb.2020.04.006.
[152] Feng Z, Chen M, Liang T, Shen M, Chen H, Xie XQ. Virus-CKB: an integrated bioinformatics platform and analysis resource for COVID-19 research. Brief Bioinforma 2021;22(2):882–95. https://doi.org/10.1093/bib/bbaa155.