Computational and Structural Biotechnology Journal 20 (2022) 5713–5728

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

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

Recent computational drug repositioning strategies against SARS-CoV-2

Lu Lu [a,b,1], Jiale Qin [a,c,1], Jiandong Chen [a,d], Na Yu [a], Satoru Miyano [e], Zhenzhong Deng [f,*], Chen Li [a,b,g,*]

[a] Department of Human Genetics, Department of Ultrasound, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China
[b] Zhejiang Provincial Key Laboratory of Genetic & Developmental Disorders, Zhejiang University School of Medicine, Hangzhou, China
[c] Zhejiang Provincial Key Laboratory of Precision Diagnosis and Therapy for Major Gynecological Diseases, Hangzhou, China
[d] School of Public Health, Undergraduate School of Zhejiang University, Hangzhou, China
[e] M&D Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan
[f] Xinhua Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China
[g] Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou, China

ARTICLE INFO

Article history:
Received 5 May 2022
Received in revised form 12 October 2022
Accepted 12 October 2022
Available online 17 October 2022

Keywords:
COVID-19
Drug repositioning
Drug combination
Neural network
Signature matching
Molecular docking

ABSTRACT

Since COVID-19 emerged in 2019, significant levels of suffering and disruption have been caused on a global scale. Although vaccines have become widely used, the virus has shown its potential for evading immunities or acquiring other novel characteristics. Whether current drug treatments are still effective for people infected with Omicron remains unclear. Due to the long development cycles and high expense requirements of de novo drug development, many researchers have turned to consider drug repositioning in the search to find effective treatments for COVID-19. Here, we review such drug repositioning and combination efforts towards providing better handling. For potential drugs under consideration, aspects of both structure and function require attention, with specific categories of sequence, expression, structure, and interaction, the key parameters for investigation. For different data types, we show the corresponding differing drug repositioning methods that have been exploited. As incorporating drug combinations can increase therapeutic efficacy and reduce toxicity, we also review computational strategies to reveal drug combination potential. Taken together, we found that graph theory and neural network were the most used strategy with high potential towards drug repositioning for COVID-19. Integrating different levels of data may further improve the success rate of drug repositioning.

1. Introduction

Beyond its first discovery in 2019, COVID-19 has become a global pandemic. It is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [1]. People with COVID-19 exhibit a range of symptoms ranging from mild to severe. For the latter, acute respiratory distress syndrome (ARDS) is of particular concern. As triggered by nonspecific inflammatory cell infiltration and a local cytokine storm, it is characterized by difficulty in breathing and low blood oxygen levels, potentially leading to respiratory failure. ARDS is listed as the cause of death for 70 % of directly fatal COVID-19 cases. In addition to attacking the lungs, COVID-19 related uncontrolled inflammation can also inflict multi-organ damage, especially cardiac, hepatic, and renal systems, which incorporate the primary causes of death for nearly all (28 %) of the remaining 30 % of directly fatal cases [2].

To consider drugs that may target the virus and/or disrupt its various stages of invasion and propagation within the body, it is important here to give a brief overview of the SARS-CoV-2 virus itself and its invasive strategy. The SARS-CoV-2 consists of the genome and the membrane that envelops it. It exists in a roughly spherical shape where prominent club-shaped surface projections represent its spike protein (S protein) and give it its characteristic 'crown-like' appearance [3,4]. These spikes are tools for the virus to invade the human body. When the spikes contact the target cell, the S protein will attach to angiotensin-converting enzyme 2 (ACE2, an import modulator involved in blood pressure regulation

Abbreviations: DEGs, differentially expressed genes; AA, amino acids; DNN, Deep Neural Network; AI, artificial intelligence; PPI, protein-protein interactions; KG, knowledge graph; GCN, Graph Convolutional Network; GEP, gene expression profiles.

* Corresponding authors at: Department of Human Genetics, Department of Ultrasound, Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China (C. Li).

E-mail addresses: dengzhenzhong@xinhuamed.com.cn (Z. Deng), chenli2012@zju.edu.cn (C. Li).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2022.10.017

[5]) on the respiratory tract cell surface, triggering endocytosis of SARS-CoV-2 virus and the formation of an endosome. Endosome encapsulation therefore enables the virus to escape from the immune system. Within the endosome, the S protein can then be cleaved by the proprotein convertase furin at the S1/S2 site, and the transmembrane serine protease 2 (TMPRSS2) at the S2′ site, thus enabling the protein to refold and change its conformation. The newly exposed part then fuses with the endosome membrane for subsequent release [6]. Virus RNA is then able to be translated into a polypeptide to synthesize RdRp under the action of 3CL protein. RdRp is RNA polymerase that allows RNA to replicate in the body. Concurrently, as the virus genome begins to replicate, the invaded host cell is now active in the production of virus proteins instead of cellular proteins. These newly produced proteins and viral genome are soon assembled to form the new virus which then escapes from the host cell to invade other cells [7,8] (Fig. 1).

On November 26, 2021, the WHO designated Omicron (B.1.1.529) as the fifth "Variant of Concern" (VOC) after the Alpha, Beta, Gamma, and Delta Variants. Delta has 13 mutations, of which nine are on the spike protein and two are on the receptor-binding domain. Compared to Delta, Omicron is by far the most heavily mutated variant. It has 50 mutations overall, with at least 32 variants on the spike protein, 10 of which are on the receptor-binding domain. Omicron is highly contagious with only a three-day incubation period. Although vaccines are becoming popular, the ability of Omicron to facilitate immune escape seems to be stronger than that of previous variants, with existing infected and vaccinated people increasingly likely to become reinfected. The process of drug discovery for SARS-CoV-2 therefore must aim to be as adaptive as the virus itself, and to evolve continually to keep up with the virus' own evolution. A comparison of the features of Omicron and other VOCs is shown in Table 1.

Because of the high probability of mutation of the SARS-CoV-2 virus, researchers are focusing on the development of antiviral agents against the more conserved proteins among multiple coronaviruses [7,9–11]. Conserved proteins include RdRp, and the main protease (Mpro). Remdesivir, authorized for emergency use by the Food and Drug Administration (FDA), works by limiting a virus' ability to replicate itself within the body. It has shown to have antiviral activity against SARS-CoV-2 *in vivo* in rhesus monkeys through its targeting of RdRp. However, one other clinical study initially concluded that Remdesivir had no statistically significant clinical benefits for severe COVID-19 patients [12]. Other studies have viewed host protease as a drug target, including the targeting of the cell surface proteases TMPRSS2 and furin. Bromehexine is an oral mucolytic that has been identified as a TMPRSS2 inhibitor in a high-throughput screening study. Its former use is to primarily treat prostate cancer. However, in a recent biochemical study it

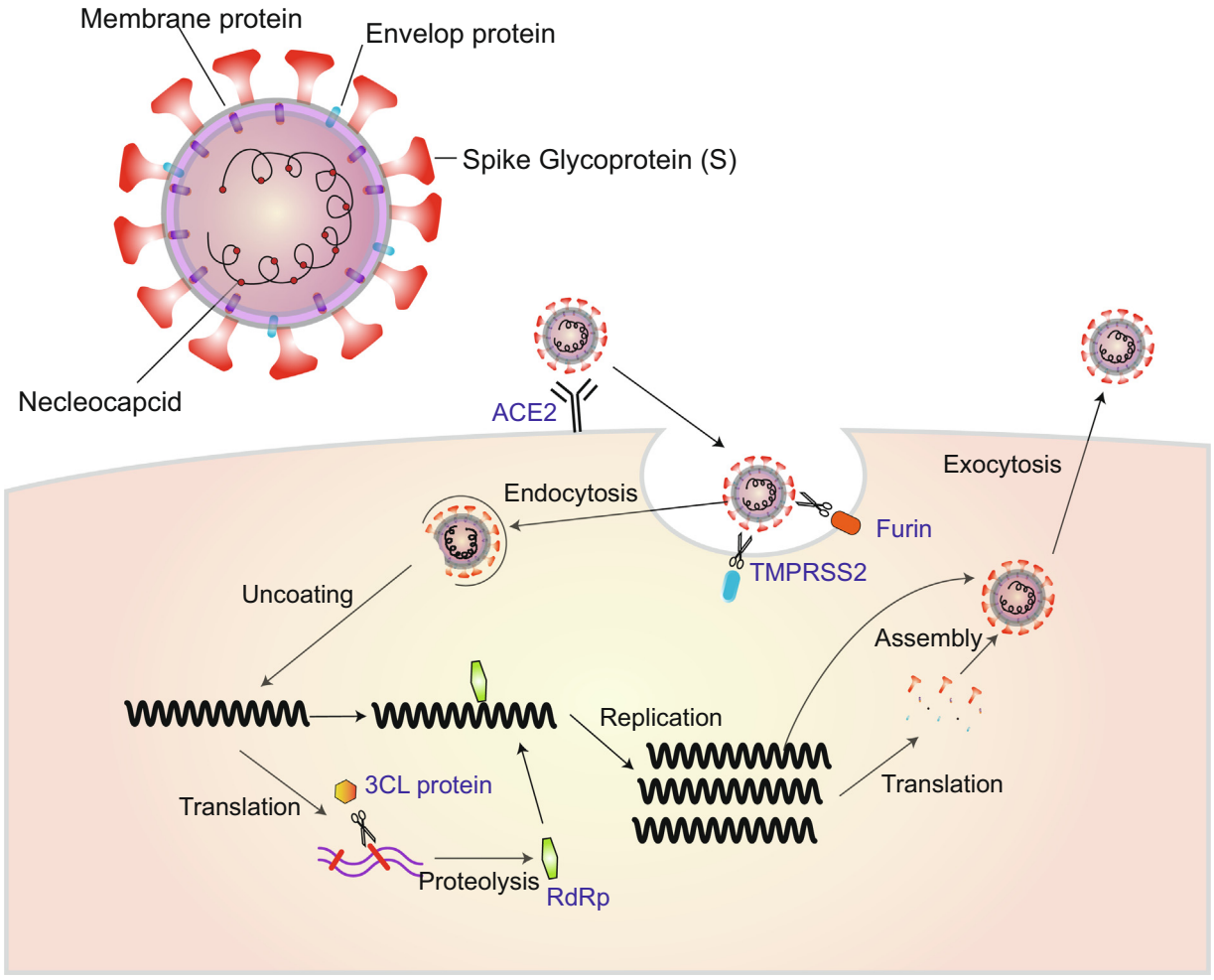


Fig. 1. The virus invasion of the human host cell. Following SARS-Cov-2 entering the cell, SARS-CoV-2 binds to the cell receptor ACE2, then promotes cell membrane fusion under the action of TMPRSS2 and Furin. The virus RNA is translated into a polypeptide to synthesize RdRp under the action of 3CL protein. RdRp is the RNA polymerase which allows RNA to replicate in the body. The replicated RNA and translated protein reassemble into a new virus. This new virus is released outside the cell through exocytosis to further infect other cells.

Table 1
The comparison of different VOC of SARS-CoV-2.

Variants of concern	Alpha	Beta	Gamma	Delta	Omicron
Lineage	B.1.1.7	B.1.351	P.1 (B.1.1.28.1)	B.1.617.2	B.1.1.529
Emergence	First detected in United Kingdom in November 2020	First detected in South Africa in October 2020	First detected in January 2021 and tracked to Brazil	First identified in India in December 2020	Emerged in South Africa's Gauteng province in mid-October 2021
Mutations	One or more mutations in the virus' spike protein	Three mutations of particular interest in the spike region	10 defining mutations in its spike protein	Overall 13 mutations: 9 on spike protein and 2 on RBD	$\geq$ 32 variants on spike protein and 10 on RBD
Transmission	40–80 % more transmissible than wild-type SARS-CoV-2 by estimation	–	About 2.0 times (50 % CrI, 1.7–2.4 times) more transmissible	One person infects about six people, more than original virus (original- two or three people)	High transmissibility
Incubation	Six days on average	–	–	Four-day incubation period; faster than the original virus (original- six days on average).	Three-day incubation
Immune escape	Blunts the potency of infection-blocking 'neutralizing' antibodies	More resistant to immunity generated by vaccines or previous infections than are other variants, including Delta	Evading about 32 % (50 % CrI, 21–46 %) of inherited immunity from previous coronavirus diseases, leading to the possibility of reinfection	Antibodies created by older strains remain effective	Escapes the majority of existing SARS-CoV-2 neutralizing antibodies
Severity of Illness	May be associated with a higher degree of mortality, (awaiting further evidence for confirmation)	People infected with Beta were 25 % more likely than those infected with Alpha to develop severe disease, as well as 57 % more likely to die	Greater chance of death than for B.1.1.28 infections	People in UK with Delta had double the hospitalization risk of those with of an earlier variant	Compared with Delta virus, Omicron patients have lower hospitalization rate, ICU hospitalization rate, and machine oxygen absorption requirement rates
Reference	[93]	[94 95]	[93]	[96]	[96,97]

Abbreviation: CrI: Credible Interval; RBD: Receptor binding domain.

failed to inhibit TMPRSS2 [9]. These studies may lead to an increased urgency to identify more specific and/or novel inhibitors against the rapid viral evolution of SARS-CoV-2 variants in human hosts.

Traditional drug development is a difficult, expensive, and time-consuming process. On average it takes 10 $\sim$ 15 years and $1.5 to $2.6 billion to bring a new drug to market [13]. Drug repurposing provides an effective alternative method for higher-speed, lower-risk, and lower cost drug development. It aims to discover new use for existing or abandoned drugs. These drugs have often previously undergone Phase I of clinical trial, have already passed certain safety requirements, and have often already been licensed for human use. These factors considerably accelerate the drug's developmental timelines and expense for repurposing by about 3–6 years and about $300 million, respectively [14].

Alsharif et al [15–17] has reviewed the potential for artificial intelligence technology to be employed specifically to counter the COVID-19 virus in three main areas: 1) in rapid diagnosis and detection using X-ray and CT scans; 2) in prediction of outbreak virus spread; and 3) in the search for potential treatments. Santamaría et al [18], similarly discussed the integration of heterogeneous biomedical data to aid drug repositioning, suggesting five different paths in the search repurposable drugs: 1) COVID – symptoms – drugs; 2)COVID – symptoms – diseases – drugs; 3) COVID – symptoms – diseases – genes – targets – drugs; 4) COVID – genes – diseases – drugs; and 5) COVID – genes – targets – drugs. Mule et al [19] has also reviewed the biological targets related to SARS-COV-2, including targets associated with the virus and the targets associated with the drug. To compliment, rather than repeat, their commendable efforts, here we incorporate a small molecule focus, and consider not only artificial intelligence technology or target-based methods, but also other computational methods.

This study therefore aims to not just discuss the existing drug repositioning methods based on recent studies, but also to inspire new drug repositioning methods for COVID-19. The key contributions of this work are summarized as follows:

1) Categorizing available data into four groups: sequence data, expression data, structure data and interaction data;

2) Discussing the available methods for each aspect of the given data;

3) Highlighting the potential for combination therapies to play essential roles in antiviral therapies, often providing improved efficacy and reduced toxicity [20], we therefore explore methods for drug combination discovery.

2. Available data

Since its outbreak in 2019, our understanding of COVID-19 has increased considerably relating to its virus genome [21,22], its transcriptomic and proteomic expression profiles [23–25], its protein structure [26–29], its virus-host protein interactions [30–32], and its relationship related to other coronavirus variants [33]. A bundle of related resources has also been collected and summarized by the National Center for Biotechnology Information (NCBI) (see summary below).

- Virus genome

  GenBank: NC045512, MN908947, MN938384, MN975262.

- Transcriptomic and proteomic expression profiles

GEO: GSE147507, GSE162131, GSE153970; BIG: CRA002390; NGDC: PRJCA002273; iProX: IPX0002106000, IPX0002171000, IPX0002393000; PRIDE: PXD017710.

- COVID-19 relevant protein structure

3CLpro (PDB ID: 6LU7), PLpro (PDB ID: 6WX4), RdRp (PDB ID: 7BV2, 6 M71, 7BTF), spike receptor-binding domain (RBD) (PDB ID: 6M0J), N protein (PDB ID: 6M3M) and ACE2 (PDB ID: 1R42).

- Virus-host protein interactions

332 high-confidence protein interactions between SARS-CoV-2 proteins and human proteins (https://public.ndexbio.org/#/network/43803262-6d69-11ea-bfdc-0ac135e8bacf).

- COVID-19 registry

https://clinicaltrials.gov/ct2/results?cond = COVID-19.

- Other coronavirus variants:

SARS-CoV (GEO: GSE1739, GSE33267, AY390556, AY485277, AY508724, AY278489), MERS-CoV (GEO: GSE122876, KT006149, KM027262), HCoV-229E (GEO: MN306046) and HCoV-NL63 (GEO: MG77280.

- Bundled resource related to SARS-CoV-2

https://www.ncbi.nlm.nih.gov/sars-cov-2/.

The related knowledge of existing drugs including chemical structures [34–37], drug-target interactions [34–36,38–42], drug perturbations [43,44], phenotype effect [36,40,45–47] and drug classifications [48,49] has been identified and accumulated here. A brief summary of these related drug resources is provided in Table 2.

3. Computational techniques for drug repositioning

For different data types, different methods have been used to relocate relevant drugs. Fig. 2 shows a schematic overview of the available data sets and corresponding methods. On the left differ-

Table 2
Drug-related information.

Resource	Type	Description	URL
ChEMBL	General database	Manually curated database of bioactive molecules	https://www.ebi.ac.uk/chembl/
DrugCentral	General database	Drug information resource	https://drugcentral.org/
PubChem	General database	Database of chemical structures, identifiers, related diseases of molecules	https://pubchem.ncbi.nlm.nih.gov/
ZINC15	Chemical 3D structure	Database of 3D formats of compounds	https://zinc15.docking.org/
DrugBank	Drug–target associations	Database of drugs and drug targets	https://go.drugbank.com/
STITCH	Drug-target associations	Interaction networks of chemicals and proteins	https://stitch.embl.de/
Target Therapeutic Database	Drug–target associations	Database of therapeutic protein and nucleic acid targets, targeted disease, pathway information and the corresponding drugs directed at each of these targets	https://bidd.nus.edu.sg/group/cjttd/
BindingDB	Drug–target associations	Database of measured binding affinities between drugs and proteins	https://www.bindingdb.org/bind/index.jsp
Guide To Pharmacology	Drug–target associations	An expert-driven guide to pharmacological targets and the substances that act on them.	https://www.guidetopharmacology.org/
PharmGKB	Drug–target associations, Drug's indications	A resource assessing the impact of genetic variation on drug response	https://www.pharmgkb.org/
BindingDB	Drug–target associations	Interactions of proteins with drug-like molecules	https://www.bindingdb.org/bind/index.jsp
cMap	Drug perturbations	Collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules	https://www.lincsproject.org/
LINCS	Drug perturbations	Database of gene expression when cells are exposed to a variety of perturbing agents	https://cancergenome.nih.gov/
CTD	Chemical-disease interactions, Gene-disease interactions, Chemical-protein interactions	A database of chemical–gene/protein interactions, chemical–disease and gene–disease relationships	https://ctdbase.org/
SIDER	Drug's side effects	A database of adverse drug reactions	https://sideeffects.embl.de/
FAERS	Drug's side effects	A database of adverse drug reactions	https://open.fda.gov/data/faers/
DynaMed	Drug's indications, contraindications and adverse reactions	Database of evidence-based drug metadata.	https://www.dynamed.com/
ATC	Drug ontology	Medicinal products classified according to the main therapeutic use of their main active ingredient.	https://www.whocc.no/atc_ddd_index/
ChEBI	Drug ontology	Dictionary of molecular entities focused on 'small' chemical compounds	https://www.ebi.ac.uk/chebi/

ent data type categories, representing different levels of biological systems, are presented. On the right side the relevant computational approaches are shown. Arrows link the different levels of data to their corresponding methods. For example, for expression data, signature mapping and neural network can be employed for drug repositioning. We will now go on to discuss the computational techniques that have been successfully exploited for each level of data. Whilst some methods may cross multiple levels of data, here we simply focus upon the most important or relevant level of data in our explanations.

3.1. Sequence-based computational techniques

The genome sequence is the basis of viral inheritance. The genome of SARS-CoV-2 contains a positive-sense and single-stranded RNA of about 30 kb size [4]. Since its outbreak, the COVID-19 genome of different regions has been successfully sequenced over time. The viral nucleotide sequence can be incorporated into a metabolic network with nodes (representing chemical compounds or metabolites) and edges (identifying reaction that can be catalyzed by one or more enzymes). In this representation, it is noted that an excessive concentration of a compound, that has accumulated as result of a particular enzyme, can result in a particular observed pathology. Thus, these enzymes can be considered as targets for possible therapies [50]. Here flux balance analysis (FBA), can be incorporated as a mathematical approach for analyzing the flow of metabolites in such a metabolic network [51].

Renz et al. firstly converted the genome and protein sequences to counts of nucleotides and amino acids. Then, by incorporating these nucleotides, amino acids, Adenosine triphosphate (ATP), and the liberation of pyrophosphate (PPi) of SARS-CoV-2 into an existing infected alveolar macrophage model, they were able to build a vital biomass objective function (VBOF). By optimizing the VBOF growth rate, they found the stoichiometric and metabolic changes between uninfected and infected host cells and identified the potential antiviral targets using reaction knock-outs and host-derived enforcement approaches [52]. As a result, the authors highlighted guanylate kinase (GK1) for potential use for antiviral therapies against SARS-CoV-2.

FBA is therefore commonly used for analyzing diseases caused by pathogens. It represents a rapid repositioning method related to genome sequences. However, this study also has some limitations: i) the VBOF only considers amino acids, nucleotides, and energy requirements, and does not consider virus-host cell recognition, viral entry, or lipid envelope production or release; ii) it uses the genome-scale metabolic model (*GEM*) of human alveolar macrophages, rather than that of airway epithelial cells (AEC).

3.2. Gene expression-based computational techniques

Therapeutic interventions need to consider the perturbation of disease system properties, and have less to do, functionally speaking, with genetic and genomic events alone [53]. We often use gene expression changes in the primary descriptions of perturbed disease systems. The COVID-19 expression profiles can be retrieved

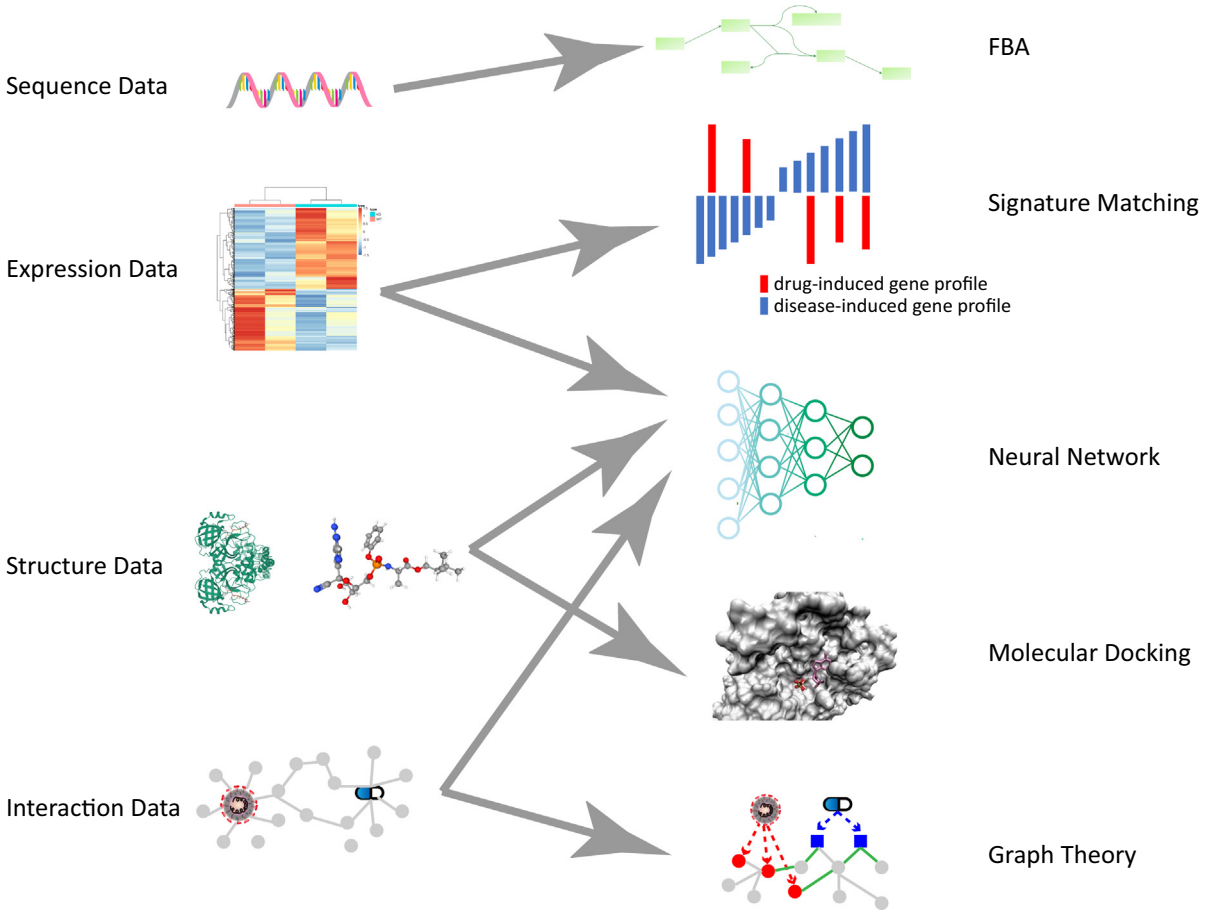


Fig. 2. The relationship between available data and corresponding methods. On the left side are the available data types, with the corresponding methods on the right. 1) For sequence data, a common approach is FBA; 2) For expression data, the common methods are signature matching and neural network; 3) For structure data, neural network and molecular docking are generally used; 4) For interaction data, neural network and graph theory can be taken. The sub-figure about molecular docking are the screenshot . adopted from https://www.youtube.com/shorts/oeqJo9xYviY

from GEO or Array Express, which contains raw gene expression data from hundreds of disease conditions in human and animal models. Meanwhile, drug perturbations can be derived from Connectivity Map (cMap) [43], which consists of gene expression profiles (GEP) generated via the dosing of ~ 1,300 compounds in five human cancer cell lines. The next generation of cMap is the L1000 platform, part of The Library of Integrated Network-Based Cellular Signatures (LINCS) [44]. The first installment of L1000 encompasses ~ 1,400,000 GEPs generated with ~ 20,000 compounds upon treatment of ~ 50 human cell lines. Based on such gene expression data, signature matching and neural network methods can be then used for drug repositioning.

- **Signature matching**

Signature matching, also known as signature reversion, involves investigation of whether disease expression patterns can be reversed on a molecular level. This has been applied specifically for COVID-19 [50], and shown to be of potentially therapeutic benefit. On one hand, COVID-19′s signature can be identified by comparing the gene expression profiles between COVID-19 patients and unaffected controls. On the other hand, the molecular signature can be evaluated by comparing the expression changes before and after treatment with small molecules. A commonly used method of signature matching is Gene Set Enrichment Analysis (GSEA) [54], a computational method that determines whether a priori defined set of genes shows statistically significant differences between two biological states. In our scenario, GSEA is used to determine whether the differentially expressed genes (DEGs) between COVID-19 patients and unaffected controls are enriched in the drug perturbed expression profile.

Mahmud et al. performed transcriptomic RNA-seq analysis of idiopathic pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD) and COVID-19, revealing 65 shared DEGs. The hub-genes of the common DEGs were viewed as drug targets, and drug molecules were identified using the Drug Signatures database (DSigDB) [55] via the interactive web tool Enrichr [56–58]. Finally, they identified 10 potential chemical compounds that have potential for repurposing against COVID-19 [59].

Furthermore, in GSEA, the DEGs between COVID-19 patients and unaffected controls can be presented along with a sign indicating whether up or down-regulation is represented, which is viewed as a query. In this case the compound molecular signature constitutes the reference database. If the up-regulated query genes appear towards the bottom of the rank-ordered molecular signature, and the down-regulated query genes appear towards the top of the rank-ordered molecular signature, this suggests that the drug can reverse the disease signature. Based on this hypothesis, Zhou et al. calculated the enrichment score (ES) for each drug and used this score to validate 135 anti-HCOV drugs. [33].

Compared to the study of Mahmud et al, Zhou et al not only considered the DEGs, but also took the sign of up or down-regulation into consideration which enables an improvement in prediction power. However, the number of perturbation data limits the application of such methods.

- **Neural network**

Neural network have developed rapidly over recent years and have facilitated achievements in natural language processing and image recognition. In the context of drug repositioning, and based on expression data, they have been used to learn the embedding representation of GEP for both COVID-19 and drugs, and then calculate the level of corresponding correlations. The most negatively related drugs with COVID-19 are then viewed as therapeutic candidates.

In this way, Pham et al. used a deep learning framework, DeepCE, to predict the GEP for novel chemicals. They took L1000 [60] experimental information, which contains the chemical compound, 978 L1000 genes, seven most frequent cell lines, and six most frequent chemical dosages, as inputs. They then transformed these into numerical representations and used them in a prediction network to predict GEP based on these representations. For drug repositing, they screened drugs in DrugBank using computations of Spearman's rank-order correlation scores of GEPs for the drug and the patient (GEO: GSE147507, NGDC: PRJCA002273) and selected the drugs that gave the most negative scores as of potentially therapeutic value. In total, they identify 10 potential drugs for PRJCA002273 and 15 drugs for GSE147507 (Table 3 for details) [61].

To summarize, changes in gene expression reflect the body's response to stimuli, such as from diseases or chemical drugs. The basic assumption for signature matching and neural network based on expression data is to seek drugs to treat diseases by reversing the phenotype of the disease. Compared with signature matching, neural network can not only utilize known drug expression profiles, but also predict expression profiles of *de novo* chemicals, which would greatly help to expand the scope for drug screening.

Although expression-based approaches are more unbiased, several drawbacks can be found. For example, if a drug or a disease does not produce a strong perturbation of gene expression, noisy profiles will be generated, leading to higher levels of false positives [50].

3.3. Structure-based computational techniques

Structure is the basis of function. The structure of COVID-19-related proteins and molecular compounds can also be used to aid in drug repositioning. Information on the structure of proteins includes the sequence of amino acids (AA) [28], secondary structure [62], and 3D structure [63]. A compound molecule also has linear representations such as SMILES [64], inChi [65], 2D structure, and 3D conformations (Table 2). Using structure-based data, neural network and molecular docking can be exploited for drug repositioning.

- **Neural network**

As for 1D structural data, nucleotide and amino acid sequences are analogous to natural language. Hence, a neural network can be employed to predict molecular-protein associations.

Ke et al. [66] used a Deep Neural Network (DNN) to identify the most important molecular descriptors from extended connectivity fingerprints (ECFPs) [67], functional-class fingerprints (FCFPs) [68], and octanol–water partition coefficient (also known as an AlogP_-count) to assign different weightings. They built an artificial intelligence (AI) platform using two independent datasets: one relating to drugs reported to be against virus, the other the known 3CL protease (See Fig. 1) inhibitors. As the infection by feline infectious peritonitis (FIP) virus in cats presented similar features to the severe acute respiratory syndrome (SARS) infection, all AI predicted drugs were then tested for activities against the feline coronavirus in an in-vitro cell-based assay. These assay results were fed back to the AI system for relearning and thus to generate a modified and improved AI model that can be subsequently reemployed to search for drug candidates. Finally, the AI system identified 80 marketed potential drugs. Among them, eight drugs showed in vitro activities against FIP, and five other drugs were also found to be active (Table 3). In this study, the authors used AI to quickly identify drugs with potential activities inhibiting SARS-CoV-2. However, their use of an in vitro cell model for feline coronavirus replication rather than SARS-CoV-2, made the results unreliable.

Table 3
Comparison of studies based on neural network.

Reference	Starting Dataset	Algorithm	Potential drugs	Description	Advantages	Limitations
Pham et al. [61]	LINCS L1000, STRING, DrugBank, COVID-19 Patient Gene Expression	DeepCE	Faldaprevir, Alisporivir, NIM811, Ceftobiprole medocaril, **Anidulafungin**, Oteseconazole, Voclosporin, **Cyclosporine**, Valspodar, Evacetrapib for PRJCA002273 and Elbasvir, Zibrentasvir, Velpatasvir, Ruzasvir, Samatasvir, Odalasvir, Coblopasvir, Baloxavir Marboxil, Metocurine, Dactinomycine, Laniquidar, Tadalaf1, SD146, AMG-487 GE-2270A for GSE147507	Models chemical substructure–gene and gene–gene associations for predicting the DEG profile perturbed by de novo chemicals	Predicts chemical-induced gene expression profiles from chemical and biological objects, especially in a de novo chemical setting	Little agreement is present among the potential drugs for two different patients
Ge et al. [85]	Drug-target-disease	CoV-DTI	CVL218	An integrative drug repositioning framework including mining knowledge graphs using GCN, literature filtering, signature matching and wet experiment evaluation	An integrative pipeline studying the mechanism of action of CVL218	CVL218 is in Phase I clinical trial
Zeng et al. [86]	Global Network of Biomedical Relationships (GNBR); and DrugBank.	CoV-KGE	**Tetrandrine**, Nadide, Estradiol, Rifampicin, Idoxuridine, Sirolimus, Deferoxamine, Prednisone, Vancomycin, Zidovdine, Ampicillin, Hydrocortisone, Etoposide, Methotrexate, Cyclosporine, Indomethacin, Etodolac, Ganciclovir, Ivermectin, **Suramin**, Clofazimine, Prednisolone, Cyclic adenosine, monophosphate, Dinoprostone, Camptothecin, **Dexamethasone**, **Lopinavir**, Emetine, **Thalidomide**, Niclosamide, **Methylprednisolone**, **Ribavirin**, **Umifenovir**, Clomifene, Mefloquine, Chloroquine, Hydroxychloro quine, Bazedoxifene, Toremifene, **Azithromycin**, Melatonin	Knowledge-graph-based deep-learning methodologies including KG embedding using rotation and validation by gene set enrichment analysis	Demonstrates a powerful deep-learning methodology to prioritize existing drugs for further investigation	Not robust to noise in graph data
Ke et al. [66]	Compounds reported or proven active against SARS-CoV, SARS-CoV-2, HIV, and influenza virus; the known 3C-like protease inhibitors		Bedaquiline, Brequinar, Celecoxib, Clofazimine, Conivaptan, Gemcitabine, Tolcapone, Vismodegib, Boceprevir, Chloroquine, HomoharrinGtonine, Tilorone, Salinomycin	Uses two AI models to learn the most important descriptors of compounds	Can quickly identify drugs with potential activities inhibiting SARS-CoV-2 based on 1D structure of compounds	Uses a feline coronavirus to validate the drug activities, which is different from SARS-CoV-2 in vitro cell model
Beck et al. [69]	Amino acid sequences; SMILES representation of $\sim$ 1,000,000,000 compounds	MT-DTI	Atazanavir, Remdesivir, Kaletra, Rapamycin, Tiotropium Bromid	Deep learning-based drug-target interaction prediction model	SMILES and AA are 1D strings. It is possible to quickly apply target proteins that do not have experimentally confirmed 3D crystal structures	Does not consider the spatial conformation of drugs and proteins
Jin et al. [89]	SARS-CoV-2 DTI data; SARS-CoV-2 targets; molecular structures	ComboNet	*Remdesivir + Reserpine, Remdesivir + IQ-1S*	Consists of two components. The first a GCN that learns the representation of a molecule and the second that models target – disease association	Performs significantly better in synergy prediction accuracy than previous methods with limited drug combination training data by incorporating additional biological information	Needs to incorporate additional biological information

Bold: in clinical trial; *italic*: in vitro test.

Beck et al. then viewed AA sequences of SARS-CoV-2 and SMILES [64] representation of drugs as natural language and used Molecule Transformer-Drug Target Interaction (MT-DTI) based on deep learning to predict drug-target interactions (The target is 3C-like proteinase of SARS-CoV-2). As a result, Atazanavir, Remdesivir, and Kaletra, were all predicted to inhibit SARS-CoV-2. Rapa-

Table 4
Comparison of studies based on molecular docking.

Reference	Starting Dataset	Target	Potential drugs	Description
Jang et al. [12]	Cocrystal structure of Mpro and RdRp; cocrystal structure of compounds	Mpro (PDB ID: 6LU7); RdRp (PDB ID: 6 M71)	*Blonanserin, Emodin (targeting Mpro), Omipalisib, Hypericin, NS-3728, Tipifarnib, LGH-447 (targeting RdRp)*	All 6,218 compounds were screened against Mpro and RdRp of SARS-CoV-2 by docking simulations
Liu et al. [71]	3CLpro structure, structure of the compounds	3CLpro (PDB ID: 6LU7)	ZINC000118795962 (Itacitinib), ZINC000003775281, ZINC000028827350 (Telcagepant), ZINC000043206238 (Vidupiprant), ZINC000100472223 (Pilaralisib), ZINC000095930125 (Poziotinib), ZINC000043131420 (Fostamatinib), ZINC000022442861, ZINC000000538550 (Ziprasidone), ZINC000009212428 (Folinic Acid), ZINC000058540931 (ITX-5061)	Uses the SCAR protocol to identify possible covalent drugs targeting 3CLpro of SARS-CoV-2
Shah et al. [70]	Protein structure of COVID-19 3CLpro with co-crystallized structure, structure of 61 reported antiviral agents	Mpro (PDB ID: 5R7Y, 5R7Z, 5R80, 5R81 and 5R82)	Lopinavir, Asunaprevir, **Remdesivir**, CGP42112A, Indinavir, Ritonavir, ABT450, Marboran (Methisazone) and Galidesivir	Uses Maestro interface to perform docking targeting 3CLpro of SARS-CoV-2 and considers 9 drugs interacting with > 2 protein structures.
Li et al. [73]	Compound from ZINC15, The 3D structures of the indicated proteins	Cathepsin B (PDB ID: 1CSB), cathepsin L (PDB ID: 5MAE), TMPRSS2 (homology model; PDB ID: 5CE1)	Trapoxin B, domatinostat (4SC-202) and (targeting CarB);neratinib (HKI-272), HKI-357 and (Z)-dacomitinib (targeting CarB and CarL) ;Iodoxamide, aceneuramic acid, (S)-boceprevir and (R)-boceprevir (targeting TMPRSS2)	Uses molecular docking towards cathepsin B, cathepsin L, and TMPRSS2
Chen et al. [29]	FDA-approved drugs from ZINC15 database and Taiwan NHI-approved drugs from the website of NHI, 3CLpro, PLpro, RdRp, spike receptor-binding domain (RBD), N protein, ACE2	spike receptor-binding domain (RBD) (PDB ID: 6M0J), 3CLpro (PDB ID: 6LU7), RdRp (PDB ID: 7BV2), PLpro (PDB ID: 6WX4), N protein (PDB ID: 6M3M), ACE2 (PDB ID: 1R42), TMPRSS2 (homology model; PDB ID: 5CE1)		Uses molecular docking towards 5 virus proteins and 2 host proteins

Bold: in clinical trial; *italic*: in vitro test.

mycin and tiotropium bromide were also noted as potentially effective [69].

- **Molecular docking**

If 3D conformations of related proteins to COVID-19 and drugs are available, molecular docking can then be used to simulate their interactions (Table 4). This predicts the binding geometries as well as binding energy of the drug-target complex (https://wikimili.com/en/Docking_(molecular)#cite_note-pmid8804827-1). Related studies can be divided into three categories: 1) the targeting of SARS-CoV-2 proteins; 2) the targeting of host proteins; and 3) the targeting of both SARS-CoV-2 proteins and host proteins (Fig. 3). Detailedly speaking,

i) In SARS-CoV-2 proteins, the main protease (Mpro, also called 3Clpro, Fig. 1) is involved in post-transcriptional cleavage of essential viral polypeptides. Another protein, RNA-dependent RNA polymerase (RdRp, Fig. 1) affects the replication of the virus genome. Mpro and RdRp are both conserved, so are suggested as suitable drug targets [12].

Shah et al. performed molecular docking targeting of the 3CLpro complex using Maestro interface and identified 9 antiviral drugs out of 61 antiviral molecules [70]. Jang et al. proposed a virtual drug screening strategy comprising pre-docking filtering, docking simulation, and a post-docking filter processes, to identify drug candidates targeting Mpro and RdRp. This resulted in 15 and 23 potential repurposed drugs, respectively [12]. Based on molecular docking, Liu et al. used a computational protocol named SCAR to identify possible covalent drugs targeting 3CLpro of SARS-CoV-2 [71]. They used AutoDock Vina [72] to dock the small molecules to the substrate-binding pocket of the SARS-CoV-2 3CLpro. Eleven potential covalent inhibitors of the 3CLpro of SARS-CoV-2 were identified.

ii) Targeting host proteins provides another way to fight SARS-CoV-2. Li et al. developed covalent inhibitors using the SCAR tool for TMPRSS2 (see introduction) and CatB/L, which prime the S protein together with TMPRSS2 [73].

iii) For targeting both SARS-CoV-2 proteins and host proteins, Chen et al. [29] innovatively developed DockCoV2 (https://covirus.cc/drugs/), focusing on the use of AutoDock Vina to predict the binding affinity of FDA or Taiwan National Health Insurance (NHI) approved drugs, and seven target proteins. Of the seven target proteins, five were SARS-CoV-2 proteins: (spike protein, 3CLpro, RdRp, Papain-like protease (PLpro - regulating SARS-CoV-2 viral spread and innate immunity [74]), and nucleocapsid (N) protein); and two were host proteins: (ACE2, and TMPRSS2 [29]). Subsequently, Yu et al. [75] used a plaque reduction assay to evaluate the antiviral potency of 12 compounds from DockCoV2 [29] for multiple-targets including TMPRSS2, 3CLpro and PLpro. Their strategy revealed that tamoxifen possesses an anti-SARS-CoV-2 property owing to its inhibitory performance for multiple assays.

Targeting viral proteins has high specificity and involves less damage to the human body, while targeting human proteins may detrimentally affect certain functions of the human body as such proteins may be involved in multiple physiological processes.

In summary, neural network can be used to model 1D data of structure, such as SMILES presentation of drugs and AA sequences of protein. The advantage of this is that it can quickly target proteins that do not have experimentally confirmed 3D crystal structures. However, disadvantages also exist: i) a lot of data is required

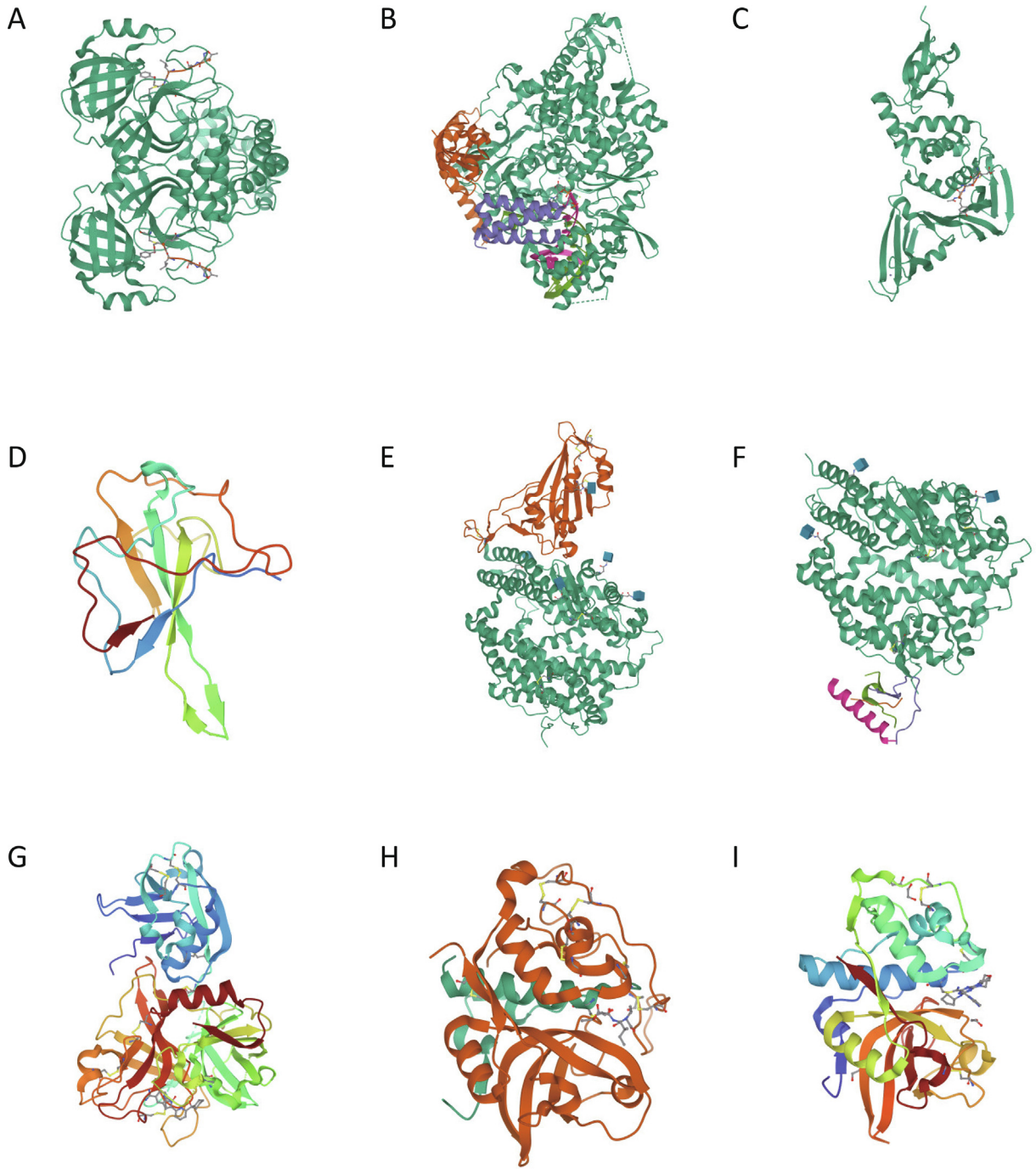


Fig. 3. Main targets of molecular docking. (A-F) SARS-CoV-2 proteins; (G-I) Host proteins. A. 3CLpro (PDB ID: 6LU7), B. RdRp (PDB ID: 7BV2), C. PLpro (PDB ID: 6WX4), D. N protein (PDB ID: 6M3M), E. spike receptor-binding domain (RBD) (PDB ID: 6M0J), F. ACE2 (PDB ID: 1R42), G. TMPRSS2 (homology model; PDB ID: 5CE1), H, Cathepsin B (PDB ID: 1CSB), and I. cathepsin L (PDB ID: 5MAE).

for the training of such a model, and ii) 1D information may not be sufficient for actual situational application. Compared to neural network, molecular docking is useful for the rapid identification of drug candidates for any known target protein, particularly if only the cocrystal structures of the target protein bounding to ligands are available [12]. However, molecular docking is deeply dependent on the cocrystal structures of the target proteins. All factors considered, structure-based approaches (both neural network and molecular docking) have proved useful and relevant to be applied to the initial stages of drug discovery.

3.4. Interaction-based computation techniques

In structure-based computational techniques, scientists model the binding of drugs and the target of proteins for drug repositioning. However, individual proteins do not operate in isolation from the complex systems, interactions, networks, and pathways incorporating many other protein players within the complexities of molecular machinery. Therefore, each drug-target interaction needs to be examined in its integrative context. Scientists have identified many interactions between viral proteins and host pro-

teins in an in-silico or experimental way [30–32]. Recent research has revealed interactions between viral RNA (vRNA) and host proteins using genome-wide CRISPR [76,77] and shRNA screening [78]. By incorporating these virus-human interactions together with human protein–protein interactions (PPI) [79–81] and drug-target interactions [82], a heterogenous network can be constructed that includes nodes as biological entities (e.g. a drug, a disease or a protein), and with edges referring to interactions. In fact, many drugs frequently show additional targets beyond the intended ones, as they are interacting with differing networks that share functional protein–protein interactions [83]. To explore the additional targets of drugs, scientists therefore have used graph theory and neural network.

- **Graph theory analysis**

The interplay between the HCoV–host interactome and drug targets in human protein–protein interaction network has been now quantified to screen for candidate drugs for HCoVs. In graph theory analysis, network proximity can be used to measure the distances between two modules, such as drug–target and disease–gene modules. Several proximity measures have been defined such as shortest, closest, separation, kernel, and centre measures. Zhou *et al.* utilized network proximity to qualify the interplay of drug targets and HCoV–host interactions in the human interactome [33]. Network proximity is measured by the average shortest path between drug targets and HCoV-associated proteins in the human protein–protein interactome. To evaluate the significance of the network distance between a drug and a given disease, they constructed a reference distance distribution by permutation tests. A z-score was used to qualify the significance of the shortest path length between targets for a drug and proteins associated with the COVID-19 modules. Finally, they prioritized 16 potential anti-HCoV potential drugs (Table 5). However, there were some limitations: i) the drugs were anti-HCoV rather than anti-*SARS-CoV-2* as the interactions between SARS-CoV-2 and host had not yet been sufficiently identified at the time they conducted their study; ii) due to lack of detailed pharmacological effects of drug targets, this study could not separate therapeutic from adverse effects; and iii) the study could not predict antiviral drugs that could targeting virus proteins directly.

Starting with a selected set of hypothesis-driven seeds (virus proteins, human proteins, or drugs), it is possible to firstly identify subnetworks connecting these seeds, then subsequently identify drug repurposing candidates associated with these mechanisms. Based on this, Sadegh et al. developed an online platform called CoVex [84]. In this, they utilized multiple graph theories (including degree centrality, closeness centrality, betweenness centrality, TrustRank, Multi-Steiner, KeyPathwayMiner) to find a subgraph of minimum cost connecting a given a set of seed nodes. Such seed nodes can be viral proteins, proteins of interest, or drugs of interest. For example, the drugs targeting viral proteins to interrupt the viral life cycle progression might be discovered from a multi-Steiner tree computation. Given a list of user-selected human host proteins, viral proteins, or drugs (referred to as seeds), users can (i) search the human interactome for viable drug targets; and (ii) identify potential drug candidates. An additional contribution to this study given by these researchers is their development of a freely accessible web server to predict drugs that provides the opportunity for users to start with their own selected proteins.

- **Neural network**

Neural network operates on a knowledge graph (KG), which contains relationships between different kinds of medical entities (e.g., diseases, drugs, and proteins). By doing so they try to predict new links between existing approved drugs and diseases. Graph Neural Network (GNN) refers to the general term for models applied by neural network to graph. They utilize structural information to predict missing links in the KG. Each node in the graph continually changes its state due to the influence of its neighbors and of further nodes until the final equilibrium is reached. The closer the neighbors are, the greater the influence. GNNs have been widely used in recommender systems, traffic prediction, computer vision, natural language processing, and in other fields.

Ge et al. [85] used GCN (Graph Convolutional Network, a kind of GNN)-based CoV-KGE for the initial screening of drugs for COVID-19. GCN updated the hidden state of all nodes iteratively to produce useful feature representations of the nodes based on an aggregate of their neighborhoods. After calculation, they derived a confidence score for each virus-target-drug pair and obtained the corresponding P-values by z-test. The drugs with P-value < 0.05 were selected as the drug candidates for each virus protein. Following the web-lab validation, they identified poly-ADP-ribose polymerase 1 (PARP1) inhibitor, CVL218, as a potentially effective drug which is currently at Phase I of its clinical trial [86]. Whilst this study provides a valuable integrative pipeline for drug repositioning, one limitation has remained in that it is only able to target virus proteins.

Zeng et al. [86] constructed a KG derived from a combination of 24 million inputs from PubMed publications or DrugBank, including 15 million edges across 39 types of relationships connecting drugs, diseases, genes, anatomies, pharmacologic classes, and gene/protein expressions, etc. They then utilized a graph embedding model, RotatE, to predict missing links between drug and HCoV-related genes. Followed by the validation in three gene expression data sets of SARS-CoV-1-infected human cells and one proteomics data set of SARS-CoV-2 infected human cells, they finally identified 41 high-confidence drug candidates for repurposing (Table 3).

Both graph theory and neural network systematically explore the entire biological network. Graph theory utilizes network proximity to seek drugs targeting disease modules. However, the real interactions between biological entities in biological network may not be as ideal as network proximity of graph theory, and at the same time, interaction networks in disease states may differ from those in normal physiological states. By contrast, neural network remains an effective way to predict missing links between drugs and targets. Despite this, the black box of neural network makes the network less interpretable. Therefore, both techniques suffer from significant limitations with the lack of experimentally validated examples of negative drug targets often resulting in many false-positive predictions.

3.5. Integrated approaches

The previously described methods repurpose drugs based on different types of data. However, each type of data represents only a partial vision of a biological system. For example, PPI networks identify potential interactions between proteins, but do not capture responses to stimuli; expression data can accurately capture stimulus-induced responses, but is of less use to extract potential interactions from them due to the problem of noise [50]; structure data considers well the interaction of drug and a single protein, but lacks any incorporation of an integrative context. For this reason, the integration of these heterogeneous data types is necessary to build a systematic view of a biological system that incorporates multiple angles for consideration and result in more accurate predictions.

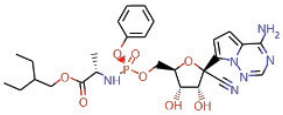
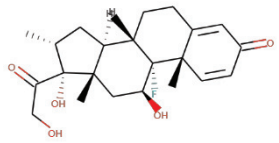
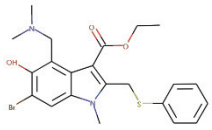
In this way, Tomazou et al. proposed a protocol to multiplex drug repositioning against COVID-19 based on multi-omics data, including expression-based data (by signature matching),

Table 5
Comparison of studies based on graph theory.

Reference	Graph	Algorithm	Potential drugs	Description	Advantages	Limitations
Zhou et al. [33]	Drug-target-disease	Network Proximity	Irbesartan, Toremifene, Camphor, Equilin, Mesalazine, Mercaptopurine, Paroxetine, Sirolimus, Carvedilol, Colchicine, Dactinomycin, Melatonin, Quinacrine, Eplerenone, Emodin, Oxymetholone, Sirolimus + dactinomycin, Mercaptopurine + Melatonin, Toremifene + Emodin	Measures the network proximity of drug targets and HCoV–host proteins in the human interactome	Systematically identifies repurposable drugs by specifically targeting HCoV-host proteins and identifying drug combinations by complementary exposure	1. Uses HCoV-host PPIs rather than SARS-CoV-2-host PPIs; 2. Cannot separate therapeutic and adverse effects; 3. Cannot predict drugs that target virus proteins directly
Sadegh et al. [84]	Drug-target-disease	Degree centrality, Closeness centrality; Betweenness centrality; TrustRank; Multi-Steiner; KeyPathwayMiner		Aims to find a subgraph of minimum cost connecting a given set of seed nodes	It offers an interactive online platform for SARS-CoV-2 host interactome exploration and drug (target) identification;	1. Virus-host interactions are still incomplete; 2. Only includes FDA approved drugs
Zhou et al. [88]	Drug-target-disease	Network proximity; propensity score (PS) matching	**Melatonin, Carvedilol**	Measures the network proximity of drug targets and HCoV–host proteins in a global interactome map.	Builds a global interactome map for SARS-CoV-2	1. Dataset remains incomplete; 2. Patient data analysis is retrospective, may have selection bias; 3. Limited for commonly used drugs due to patient data availability
Tomazou et al. [87]	Drug-target-disease	Signature matching; GWAS; taxonomy-based distances	**Dexamethasone, Beta-Estradiol, Atorvastatin, Cyclosporin A, Remdesivir, Imatinib, Hydroxychloroquine, Dactolisib, Ofloxacin, Leflunomide, Simvastatin, Pioglitazone, Methotrexate**, Cytarabine + Saracatinib, Dactolisib + Methotrexate, Hydroquinone + Vorinostat	A network-based integration of multi-omic data to prioritize the most important genes related to COVID-19 and subsequently re-rank the identified candidate drugs	Proposed drug list not only comprises drugs aiming to reverse COVID-19-induced perturbations, but also compounds with direct antiviral activity; Several of these drugs are already in clinical trials	1. Lack of harmonization across selection criteria applied for the DEGs across several datasets; 2. Selection biases among drugs might exist as observed in their null model analysis
Cheng et al. [90]	Drug-target-disease	Complementary exposure pattern	Melatonin + Toremifene	Integrates network proximity and GSEA for drug repositioning and discovery of drug combinations by complementary exposure pattern	Predicts drug combination using complementary exposure patterns	1. Interaction data is incomplete; 2. DEP (Differentially Expressed Proteins) and DEGs differ significantly due to different cell types
Renz et al. [52]	Metabolities	FBA	GK1	An integrated host-virus genome-scale metabolic model (*GEM*) of human alveolar macrophages and SARS-CoV-2	Supplies a rapid repositioning method as it can be conducted with the genome sequence	1. Does not consider virus-host cell recognition, viral entry, or lipid envelope production or release; 2. Uses a *GEM* of Human alveolar macrophages rather than epithelial cells and airway epithelial cells (AEC)

Bold: in clinical trial; *italic*: in vitro test.

Table 6
Repositioned drugs for current clinical use.

Drug name	Molecular structure	Approvals and indications	Class	Approaches	References
Remdesivir		Remdesivir is approved for the treatment of COVID-19 in adults and pediatric patients (28 days of age and older and weighing at least 3 kg) with positive results of direct SARS-CoV-2 viral testing, who are: hospitalized, or not hospitalized and have mild-to-moderate COVID-19 and are at high risk for progression to severe COVID-19, including hospitalization or death.	RdRp inhibitor	Neural network (based on structure); Integrated approach;Neural network (based on interaction data) ; Molecular docking	Beck et al. [69] Tomazou et al. [87] Jin et al. [89] Shah et al. [70]
dexamethasone		An anti-inflammatory 9-fluoro-glucocorticoid. The NIH COVID-19 Treatment Guidelines Panel recommends the use of dexamethasone in patients with COVID-19 who are receiving mechanical ventilation or in those who require supplemental oxygen but are not on mechanical ventilation. It is not recommended the use of dexamethasone or other corticosteroids in non-hospitalized patients with mild to moderate COVID-19 or in hospitalized patients with COVID-19 who do not require supplemental oxygen.	Immunosuppressant	Neural network (based on interaction data) Integrated approach	Zeng et al. [86] Tomazou et al. [87]
Umifenovir		Although data is limited, in vitro activity against SARS-CoV-1 and SARS-CoV-2 has been reported. The drug has been included in COVID-19 treatment guidelines used in China and Russia.	Antiviral agent.	Neural network (based on interaction data)	Zeng et al. [86]

phenotype-based data (by GWAS), and network-based data (measuring taxonomic distance) [87]. Through this multi-omics data integration, together with processes of drug re-ranking and drug filtering, they were able to identify a number of recently proposed drugs (including dexamethasone and remdesivir); inhibitors of Src tyrosine kinase (osutinib, dasatinib, cytarabine, and saracatinib); specific immunomodulators and anti-inflammatory drugs (dactolisib and methotrexate); and inhibitors of histone deacetylase (hydroquinone and vorinostat). Similarly, Zhou et al. build a global interactome map for SARS-CoV-2, including data of transcriptome, proteome, human interactome, and a COVID-19 registry. They then used network proximity measurement to evaluate the drug's connectivity and closeness with SARS-CoV-2 host proteins. By using propensity score (PS) matching, a series of retrospective case-control studies were conducted to test the drug–outcome relationships for COVID-19. This resulted in melatonin usage (OR = 0.48, 95 % CI 0.31–0.75) being associated with a 52 % reduced likelihood of a positive laboratory test in African Americans [88]. Such results will be useful for further academic studies and towards the development of effective medicines for the treatment of COVID-19.

Table 6 lists out the in silico repositioned drugs which have been used in a clinical setting. From the table, we can see that neural network has played a particularly important role in drug repositioning, either based on interaction data or structure data. More importantly, integrated approaches covering different levels of biological systems are shown to significantly improve prediction performance.

4. Drug combinations for COVID-19

In many disease cases, but perhaps with HIV as the most notable example, combination therapies are often more effective and/or less toxic than single drugs [89]. However, our ability to identify and validate effective combinations is limited by a combinatorial explosion of rapidly increased complexity when attempting large number of drug pairs and dosage combinations. Computational approach methods can be employed to counter this issue and what follows are two examples of computational approaches based on interaction data for drug combinations that have been applied to COVID-19.

4.1. Interaction-based computational techniques

- **Network-based approach**

Zhou et al. developed a network-based approach to identify potential drug combinations for COVID-19 based on interaction data. They found that a drug combination was therapeutically effective only if it could be both captured by the complementary exposure pattern (i.e. the targets of a drug combination hit the disease module made of SARS-CoV-2 host genes/proteins but target separate neighborhoods). In their research, they identified three potential drug combinations (Sirolimus plus Dactinomycin, Toremifene plus Emodin, and Mercaptopurine plus Melatonin) (see Fig. 4)for COVID-19 [20,33]. Cheng et al. used the same methods and discovered the combination of anti-inflammatory (Melatonin) and antiviral (Toremifene) drugs (see Fig. 4) to rescue pulmonary and cardiovascular conditions [90].

- **Neural network**

Due to the lack of high-quality training data of drug combinations, Jin et al. proposed neural network architecture that jointly learns drug – target interaction and drug – drug synergy. Firstly, they predicted the antiviral effect using a neural network based on a Drug Target Interactions (DTI) network and target–disease

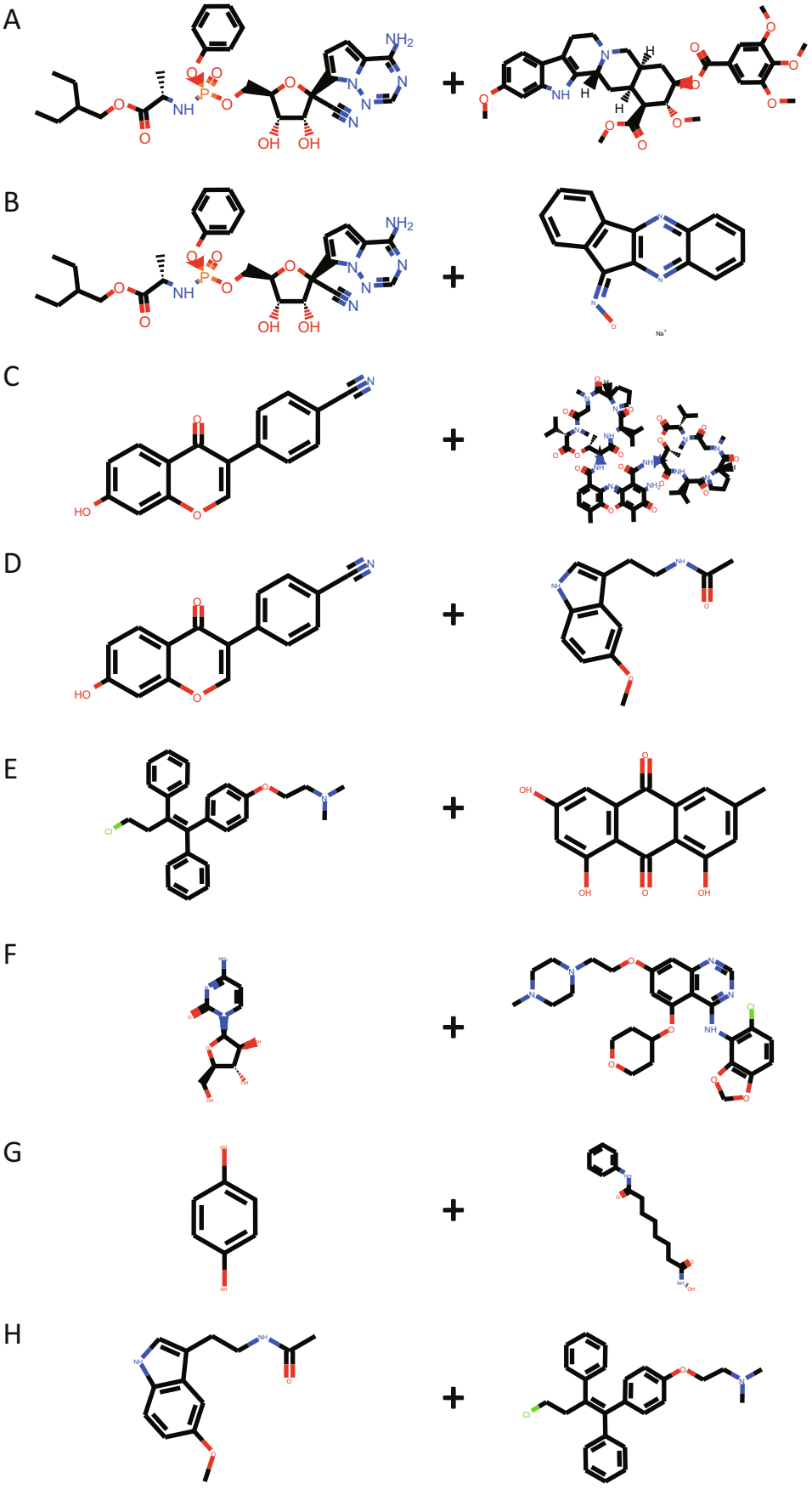


Fig. 4. Predicted drug combinations. A. Remdesivir plus Reserpine, B. Remdesivir plus IQ-1S, C. Sirolimus plus Sactinomycin, D. Mercaptopurine plus Melatonin, E. Toremifene plus Emodin, F. Cytarabine plus Saracatinib, G. Hydroquinone plus Vorinostat, and H. Melatonin plus Toremifene.

association network. They then took a list of antiviral effects of drug combinations as inputs and their synergy information as labels (synergistic or non-synergistic) to train a ComboNet model. They finally discovered two drug combinations, i) remdesivir and reserpine, and ii) remdesivir and IQ-1S (see Fig. 4), for COVID-19 [89].

5. Conclusion

Here, we have reviewed the available public data and the existing computational methods for drug repositioning and drug combinations for COVID-19. Disease-relevant data ranges from the virus genome sequence to the transcriptomic and proteomic expressions of patients; from the structure of COVID-19 related proteins to virus-host protein interactions; and from the COVID-19 registry to those of other related coronaviruses. The available drug data includes drug chemical structures, drug-target interactions, phenotype effects, and classification. Many computational methods have been implemented for drug repositioning based on such data. These include network-based methods, neural network methods, molecular docking, and signature matching. Neural network has been one of the most widely used methods across the expression data, structure data, interaction data, and have played an important role in discovering repositionable drugs for clinical use. For example, Remdesivir, as identified by MT-DTI, is a FDA-approved drug repurposed for mild-to-moderate COVID-19. Umifenovir, as predicted by CoV-KGE, has been included in COVID-19 treatment guidelines used in China and Russia. With the availability of big data, including biological, clinical, and open data (scientific publications and databases), the application of neural network will likely become more and more extensive in the future. Additionally, integrating different levels of data generates more effective drug relocation options. For example, dexamethasone, as identified by integrated tools developed by Tomazou et al., is now indicated for patients who are receiving mechanical ventilation or require supplemental oxygen. In addition, personalized medicine based on patient-specific expression data may also be an increasing trend into the future. In addition to the small molecule approach discussed in our paper, monoclonal antibodies and vaccines are two other sharp swords against the SARS-CoV-2, each case reinforcing and covering deficiencies in the other. Monoclonal antibodies, for example, can make up for the deficiencies faced by vaccines and small molecules can be used in immune-vulnerable populations. Balancing these three powerful tools, and with a joint global effort, we hope that COVID-19 will be soon overcome.

It is worth noting that there are some limitations of computational drug repositioning as discussed in this review. Firstly, we were unable to directly compare each method because they use different types of data. Secondly, repurposed drugs also require validation experiments, some being time consuming, and are not automatically or immediately available for application. Third, the tremendous volume and fast pace of published literature on the treatment of COVID-19 means that research findings and recommendations are constantly evolving as new evidence arises. Finally, precision medicine is a future trend that exists in addition to drug repositing and drug combination, for COVID-19. It may be possible to facilitate precision drug development for each person by testing the drug responses of different genotypes. With the dramatic spread of Omicron worldwide, similar methods can also be applied to the drug repositioning for Omicron. For example, according to the known sequence of Omicron (GISAID: EPI_ISL_6640916) [91], the FBA method can be used for drug repositioning. With the analysis of cryo-EM structure of spike protein–ACE2 complex (PDB ID 7T9J, PDB ID 7T9K, PDB ID 7T9L) [92], molecular docking can be employed for drug repositioning. This review also highlights the possibility of incorporating such responses into any future pandemics by perfecting the techniques of the mining of virus sequence data, expression profiles, structures, and data interaction. Overall, computational methods for drug repositioning will be highly likely to continue to provide important guidance towards epidemic disease responses.

Funding

This work was supported by grants from National Science Foundation Committee (82272126 and 82171939),a Joint Project of Science Research of National Health Commission and Zhejiang Province (WKJ-ZJ-2033); Zhejiang Provincial Science Foundation Committee (LY22H180009) and the Interdisciplinary Program of Shanghai Jiao Tong University (project number YG2019QNA40). C. Li is supported by Alibaba Cloud.

CRediT authorship contribution statement

Lu Lu: Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Jiale Qin:** Data curation, Writing – original draft, Funding acquisition. **Jiandong Chen:** Visualization, Investigation. **Na Yu:** Data curation, Validation. **Satoru Miyano:** Writing – review & editing. **Zhenzhong Deng:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision. **Chen Li:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Chris Wood of Zhejiang University for checking and editing the English writing of this manuscript. The authors are grateful to the research groups for providing the clinical information and high-throughput data related to this review. We are also grateful for the technical support by the Core Facilities, Zhejiang University School of Medicine. We thank Dr. Hangjun Wu in the Center of Cryo-Electron Microscopy (CCEM), Zhejiang University, for his technical assistance on computer clustering. We are also deeply thankful for the computational resources provided by the Super Computer System, Human Genome Center, Institute of Medical Science, University of Tokyo, Japan.

References

[1] Vargas G, Medeiros Geraldo LH, Gedeão Salomão N, et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and glial cells: Insights and perspectives. Brain Behav Immun Health 2020;7:100127.
[2] Tay MZ, Poh CM, Rénia L, et al. The trinity of COVID-19: immunity, inflammation and intervention. Nat Rev Immunol 2020;20:363–74.
[3] Klein S, Cortese M, Winter SL, et al. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. Nat Commun 2020;11:5885.
[4] Shi Y, Wang G, Cai X-P, et al. An overview of COVID-19. J Zhejiang University Science B 2020;21:343–60.
[5] El Bairi K, Trapani D, Petrillo A, et al. Repurposing anticancer drugs for the management of COVID-19. Eur J Cancer 2020;141:40–61.
[6] Benton DJ, Wrobel AG, Xu P, et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature 2020;588:327–30.
[7] Guy RK, DiPaola RS, Romanelli F et al. Rapid repurposing of drugs for COVID-19 2020;368:829-830.
[8] Sanders JM, Monogue ML, Jodlowski TZ, et al. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020;323:1824–36.
[9] Tao K, Tzou PL, Nouhin J, et al. SARS-CoV-2. Antiviral Therapy 2021;34: e00109–21.

[10] Dotolo S, Marabotti A, Facchiano A, et al. A review on drug repurposing applicable to COVID-19. Briefings Bioinf 2020;22:726–41.
[11] Dai W, Zhang B, Jiang X-M et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease 2020;368:1331-1335.
[12] Jang WD, Jeon S, Kim S et al. Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay 2021;118:e2024302118.
[13] Harrer S, Shah P, Antony B, et al. Artificial Intelligence for Clinical Trial Design. Trends Pharmacol Sci 2019;40:577–91.
[14] Benavides-Cordoba V. Drug Repositioning for COVID-19. Colombia medica (Cali, Colombia) 2020;51:e4279–e.
[15] Alsharif MH, Alsharif YH, Yahya K, et al. Deep learning applications to combat the dissemination of COVID-19 disease: a review. Eur Rev Med Pharmacol Sci 2020;24:11455–60.
[16] Alsharif MH, Alsharif YH, Albreem MA, et al. Application of machine intelligence technology in the detection of vaccines and medicines for SARS-CoV-2. Eur Rev Med Pharmacol Sci 2020;24:11977–81.
[17] Alsharif MH, Alsharif YH, Chaudhry SA, et al. Artificial intelligence technology for diagnosing COVID-19 cases: a review of substantial issues. Eur Rev Med Pharmacol Sci 2020;24:9226–33.
[18] Prieto Santamaría L, Díaz Uzquiano M, Ugarte Carro E, et al. Integrating heterogeneous data to facilitate COVID-19 drug repurposing. Drug Discovery Today 2022;27:558–66.
[19] Mule S, Singh A, Greish K, et al. Drug repurposing strategies and key challenges for COVID-19 management. J Drug Target 2022;30:413–29.
[20] Cheng F, Kovács IA, Barabási A-L. Network-based prediction of drug combinations. Nat Commun 2019;10:1197.
[21] Benson DA, Cavanaugh M, Clark K, et al. GenBank. Nucleic Acids Res 2013;41:D36–42.
[22] Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579:265–9.
[23] Blanco-Melo D, Nilsson-Payant B, Liu W-C et al. SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. 2020.
[24] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 2012;41:D991–5.
[25] Nie X, Qian L, Sun R, et al. Multi-organ proteomic landscape of COVID-19 autopsies. Cell 2021;184:775–791.e714.
[26] Wang W, McKinnie SM, Farhan M, et al. Angiotensin-Converting Enzyme 2 Metabolizes and Partially Inactivates Pyr-Apelin-13 and Apelin-17: Physiological Effects in the Cardiovascular System. Hypertension 2016;68:365–77.
[27] Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 2020;181:271–280.e278.
[28] Consortium TU. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 2020;49:D480–9.
[29] Chen TF, Chang YC, Hsiao Y, et al. DockCoV2: a drug database against SARS-CoV-2. Nucleic Acids Res 2021;49:D1152–9.
[30] Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 2020;583:459–68.
[31] Bouhaddou M, Memon D, Meyer B et al. The Global Phosphorylation Landscape of SARS-CoV-2 Infection, Cell 2020;182:685-712.e619.
[32] Gordon DE, Hiatt J, Bouhaddou M, et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science 2020;370.
[33] Zhou Y, Hou Y, Shen J, et al. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discovery 2020;6:14.
[34] Kim S, Chen J, Cheng T, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res 2018;47:D1102–9.
[35] Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2011;40:D1100–7.
[36] Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2007;36:D901–6.
[37] Sterling T, Irwin JJ. ZINC 15 – Ligand Discovery for Everyone. J Chem Inf Model 2015;55:2324–37.
[38] Alexander SP, Ball JK, Tsoleridis TJIBGtPC. Coronavirus (CoV) proteins (version 2020.4) in the IUPHAR/BPS Guide to Pharmacology Database 2020;2020.
[39] Wang X, Sheng Y, Deng H, et al. Top-N-Targets-Balanced Recommendation Based on Attentional Sequence-to-Sequence Learning. IEEE Access 2019;7:120262–72.
[40] Whirl-Carrillo M, Huddart R, Gong L, et al. An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. Clin Pharmacol Ther 2021;110:563–72.
[41] Kuhn M, von Mering C, Campillos M, et al. STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 2008;36:D684–8.
[42] Gilson MK, Liu T, Baitaluk M, et al. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology 2016;44:D1045–53.
[43] Lamb J, Crawford ED, Peck D et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease 2006;313:1929-1935.
[44] Stathias V, Turner J, Koleti A et al. LINCS Data Portal 2.0: next generation access point for perturbation-response signatures 2020;48:D431-D439.
[45] Kuhn M, Letunic I, Jensen LJ et al. The SIDER database of drugs and side effects 2016;44:D1075-D1079.
[46] Hoffman KB, Dimbil M, Erdman CB, et al. The Weber Effect and the United States Food and Drug Administration's Adverse Event Reporting System (FAERS): Analysis of Sixty-Two Drugs Approved from 2006 to 2010. Drug Saf 2014;37:283–94.
[47] Keahey D, Goldgar CJTJoPAE. Evidence-based medicine resources for physician assistant faculty: DynaMed 2008;19:43-46.
[48] Miller G, Britt HJIjob-mc. A new drug classification for computer systems: the ATC extension code 1995;40:121-124.
[49] Degtyarenko K, De Matos P, Ennis M et al. ChEBI: a database and ontology for chemical entities of biological interest 2007;36:D344-D350.
[50] Alaimo S, Pulvirenti A. Network-Based Drug Repositioning: Approaches, Resources, and Research Directions. Methods Mol Biol 2019;1903:97–113.
[51] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol 2010;28:245–8.
[52] Renz A, Widerspick L, Dräger A. FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2. Bioinformatics 2020;36:i813–21.
[53] Zhou Y, Wang F, Tang J, et al. Artificial intelligence in COVID-19 drug repurposing. Lancet Digit Health 2020;2:e667–76.
[54] Subramanian A, Tamayo P, Mootha VK et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles 2005;102:15545-15550.
[55] Yoo M, Shin J, Kim J, et al. DSigDB: drug signatures database for gene set analysis. Bioinformatics 2015;31:3069–71.
[56] Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf 2013;14:128.
[57] Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90–7.
[58] Xie Z, Bailey A, Kuleshov MV et al. Gene Set Knowledge Discovery with Enrichr 2021;1:e90.
[59] Mahmud SMH, Al-Mustanjid M, Akter F et al. Bioinformatics and system biology approach to identify the influences of SARS-CoV-2 infections to idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease patients, Briefings in Bioinformatics 2021;22:bbab115.
[60] Subramanian A, Narayan R, Corsello SM et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles, Cell 2017;171:1437-1452.e1417.
[61] Pham T-H, Qiu Y, Zeng J, et al. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. Nature Machine Intelligence 2021;3:247–57.
[62] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–637.
[63] Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–42.
[64] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 1988;28:31–6.
[65] Heller S, McNaught A, Pletnev I, et al. InChI, the IUPAC international chemical identifier. J Cheminformat 2015;7(1):23.
[66] Ke YY, Peng TT, Yeh TK, et al. Artificial intelligence approach fighting COVID-19 with repurposing drugs. Biomed J 2020;43:355–62.
[67] Rogers D, Hahn M. Extended-Connectivity Fingerprints. J Chem Inf Model 2010;50:742–54.
[68] Cereto-Massagué A, Ojeda MJ, Valls C, et al. Molecular fingerprint similarity search in virtual screening. Methods 2015;71:58–63.
[69] Beck BR, Shin B, Choi Y, et al. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol J 2020;18:784–90.
[70] Shah B, Modi P, Sagar SR. In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. Life Sci 2020;252:117652.
[71] Liu S, Zheng Q, Wang Z. Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus. Bioinformatics 2020;36:3295–8.
[72] Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–91.
[73] Li Q, Wang Z, Zheng Q, et al. Potential clinical drugs as covalent inhibitors of the priming proteases of the spike protein of SARS-CoV-2. Comput Struct Biotechnol J 2020;18:2200–8.
[74] Shin D, Mukherjee R, Grewe D, et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. Nature 2020;587:657–62.
[75] Yu P-C, Huang C-H, Kuo C-J, et al. Drug Repurposing for the Identification of Compounds with Anti-SARS-CoV-2 Capability via Multiple. Targets 2022;14:176.
[76] Daniloski Z, Jordan TX, Wessels HH et al. Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells, Cell 2021;184:92-105.e116.
[77] Wei J, Alfajaro MM, DeWeirdt PC et al. Genome-wide CRISPR Screens Reveal Host Factors Critical for SARS-CoV-2 Infection, Cell 2021;184:76-91.e13.
[78] Zhang S, Huang W, Ren L, et al. Comparison of viral RNA–host protein interactomes across pathogenic RNA viruses informs rapid antiviral drug discovery for SARS-CoV-2. Cell Res 2021.
[79] Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2012;41:D808–15.
[80] Kotlyar M, Pastrello C, Malik Z, et al. IID 2018 update: context-specific physical protein–protein interactions in human, model organisms and domesticated species. Nucleic Acids Res 2018;47:D581–9.

[81] Cheng F, Desai RJ, Handy DE, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun 2018;9:2691.
[82] Cheng F, Liu C, Jiang J, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol 2012;8: e1002503.
[83] Alaimo S, Pulvirenti A. Network-Based Drug Repositioning: Approaches, Resources, and Research Directions. In: Vanhaelen Q, editor. Computational Methods for Drug Repurposing. Springer, New York: New York, NY; 2019. p. 97–113.
[84] Sadegh S, Matschinske J, Blumenthal DB, et al. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. Nat Commun 2020;11:3518.
[85] Ge Y, Tian T, Huang S, et al. An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. Signal Transduction and Targeted Therapy 2021;6:165.
[86] Zeng X, Song X, Ma T, et al. Repurpose Open Data to Discover Therapeutics for COVID-19 Using Deep Learning. J Proteome Res 2020;19:4624–36.
[87] Tomazou M, Bourdakou MM, Minadakis G, et al. Multi-omics data integration and network-based analysis drives a multiplex drug repurposing approach to a shortlist of candidate drugs against COVID-19. Briefings Bioinf 2021;22.
[88] Zhou Y, Hou Y, Shen J, et al. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. PLoS Biol 2020;18:e3000970.
[89] Jin W, Stokes JM, Eastman RT et al. Deep learning identifies synergistic drug combinations for treating COVID-19 2021;118:e2105070118.
[90] Cheng F, Rao S, Mehra R. COVID-19 treatment: Combining anti-inflammatory and antiviral therapeutics using a network-based approach. Cleve Clin J Med 2020.
[91] Kandeel M, Mohamed MEM, Abd El-Lateef HM et al. Omicron variant genome evolution and phylogenetics;n/a.
[92] Mannar D, Saville JW, Zhu X, et al. SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein-ACE2 complex. Science 2022:: eabn7760.
[93] SARS-CoV-2 Alpha variant. https://en.wikipedia.org/wiki/SARS-CoV-2_Alpha_variant.
[94] SARS-CoV-2 Beta variant. https://en.wikipedia.org/wiki/SARS-CoV-2_Beta_variant.
[95] Callaway E. Remember Beta? New data reveal variant's deadly powers.
[96] Mille K. Omicron vs. Delta COVID-19 Variants: Do We Need Another Vaccine? https://www.prevention.com/health/a38400888/omicron-vs-delta-covid-19-variant-comparison/.
[97] Cao Y, Wang J, Jian F, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. Nature 2021.