Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

AI-powered drug repurposing for developing COVID-19 treatments[*]

Zhichao Liu[a], Xi Chen[a], Wendy Carter[b], Alicia Moruf[b], Takashi E Komatsu[b], Sonia Pahwa[b], Kirk Chan-Tack[b], Kevin Snyder[b], Nicholas Petrick[c], Kenny Cha[c], Madhu Lal-Nag[d], Qais Hatim[d], Shraddha Thakkar[d], Yu Lin[e], Ruili Huang[f], Dong Wang[a], Tucker A Patterson[a], and Weida Tong[a], [a]National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR, United States; [b]Office of New Drugs, Center for Drug Evaluation and Research, US FDA, Silver Spring, MD, United States; [c]Office of Science and Engineering Labs, Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD, United States; [d]Office of Translational Sciences, Center for Drug Evaluation and Research, US FDA, Silver Spring, MD, United States; [e]National Human Genome Research Institute, National Institutes of Health, Rockville, MD, United States; [f]National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, United States

[*]*Disclaimer*: This book chapter reflects the views of the authors and does not necessarily reflect those of the U.S. Food and Drug Administration or the National Institutes of Health. Any mention of commercial products is for clarification only and is not intended as approval, endorsement, or recommendation.

Abstract

Emerging infectious diseases are an ever-present threat to public health, and COVID-19 is the most recent example. There is an urgent need to develop a robust framework to combat the disease with safe and effective therapeutic options. Compared to de novo drug discovery, drug repurposing may offer a lower-cost and faster drug discovery paradigm to explore potential treatment options of existing drugs. This chapter elucidates the advantages of artificial intelligence (AI) in enhancing the drug repurposing process from a data science perspective, using COVID-19 as an example. First, we elaborate on how AI-powered drug repurposing benefits from the accumulated data and knowledge of COVID-19 natural history and pathogenesis. Second, we summarize the pros and cons of AI-powered drug repurposing strategies to facilitate fit-for-purpose selection. Finally, we outline challenges of AI-powered drug repurposing from a regulatory perspective and suggest some potential solutions.

Key points

- AI-powered drug purposing is promising for emerging treatments for COVID-19 infection.
- Accumulated biological data profiles facilitate AI-based drug repurposing efforts for development of COVID-19 therapies.
- The 'fit-for-purpose selection of AI-powered drug repurposing strategies is key to uncovering hidden information among drugs, targets, and diseases.
- Efforts from different stakeholders boost the adoption of AI-powered drug repurposing in the regulatory setting.

Introduction

The devastating effect of the COVID-19 pandemic represents a massive global health crisis with millions of deaths and devastating social impairments (Cutler and Summers, 2020; Bavel et al., 2020). The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the highly contagious infectious disease COVID-19, has infected over 54 million people, killed over 0.82 million in the US (as of 01/02/2022), and the case number is still surging (https://covid.cdc.gov/covid-data-tracker/#datatracker-home). Tremendous efforts have been made to fight COVID-19, which encompassed almost every aspect of public

health (Parasher, 2021). Remarkably, over 1000 clinical trials have been conducted or are ongoing to explore treatment options for COVID-19 and a few encouraging milestones have been achieved. For example, to date (as of 01/02/2022) there is one fully FDA-approved vaccine (i.e., the Pfizer-BioNTech COVID-19 Vaccine) for prevention of COVID-19 and one drug (i.e., remdesivir) approved for treatment of patients hospitalized with COVID-19 (Fang et al., 2020). However, the evolving nature of SARS-CoV-2 has become one of the significant obstacles to compelling development of suitable treatments. Although most mutations in the SARS-CoV-2 genome are either deleterious and swiftly purged or relatively neutral (Harvey et al., 2021), a small proportion of mutations (e.g., Delta variant) differ in their functional properties. These mutations may enhance transmission, increase disease severity, and/or escape from host immunity (Planas et al., 2021). Therefore, continued development of effective and safe treatments for COVID-19 is still required.

Drug repurposing explores new uses for previously approved drugs, which has been demonstrated to provide a faster, safer, cost-effective paradigm as an alternative means to building a de novo drug discovery pipeline (Pushpakom et al., 2019). Consequently, many advocates are promoting COVID-19 treatment development relying upon drug repurposing approaches (Harrison, 2020). Drug repurposing has been intensively investigated and successfully implemented for various diseases as seen in oncology (Zhang et al., 2020), rare diseases (Delavan et al., 2018), and neglected tropical diseases (Klug et al., 2016). Conventional drug repurposing approaches are primarily based on close clinical observations by domain experts, as exemplified by the story of sildenafil, the discovery of which was a serendipitous or so-called "happy accident" (Gil and Martinez, 2021). In contrast, computational drug repurposing applies diverse biological profiles to comprehensively explore repurposing opportunities of approved drugs and existing compounds with a quick cycle of hypothesis generation and verification (Liu et al., 2013).

Artificial Intelligence (AI) is changing the landscape of biomedicine, creating opportunities for innovative drug development (Vamathevan et al., 2019) and reshaping public health (Panch et al., 2019) (Fig. 1). Notably, various offshoots of AI have been applied and played a vital role in tackling COVID-19 (Arora et al., 2020), including forecasting the spread of the virus (Arora et al., 2020), monitoring contact tracing (Maghdid et al., 2020), accelerating early diagnosis (Li et al., 2020), facilitating communication between patients and healthcare practitioners (Miner et al., 2020), and developing treatments (Zhou et al., 2020a; Mohapatra et al., 2020). Integration of advances in AI and drug repurposing principles may hold great promise in fighting COVID-19. More importantly, the accumulated knowledge and developed AI-powered drug repurposing frameworks could be leveraged to react to future emerging infections (Zhou et al., 2020b). Herein, we summarize the challenges and opportunities of AI-powered drug repurposing in developing COVID-19 treatments and lessons learned from a data science perspective. First, we suggest a reliable and robust data warehouse for managing the diverse biological data profiles to enable reproducible AI-powered drug repurposing development. Second, we point out the pros and cons of existing AI-powered drug repurposing frameworks to facilitate 'fit-for-purposing' selections. Finally, we set out the remaining challenges for clinical adoption of AI-powered drug repurposing in a regulatory setting and suggest potential solutions for further improvement.

A wealth of data resources on COVID-19 enables drug repurposing development

Enormous efforts around the world are being made to respond to the COVID-19 pandemic. Consequently, diverse data profiles have been generated to facilitate a better understanding of COVID-19 etiology and pathogenesis. Meanwhile, many clinical trials have been conducted and high-throughput screening data have been generated looking for potential treatment options. The accumulated data profiles cover a broad spectrum of different biological aspects, serving as a starting point for developing AI-powered drug repurposing approaches. The data resources are summarized into a COVID-19-centric (Table 1) and drug-centric (Table 2) framework based on different drug repurposing principles.
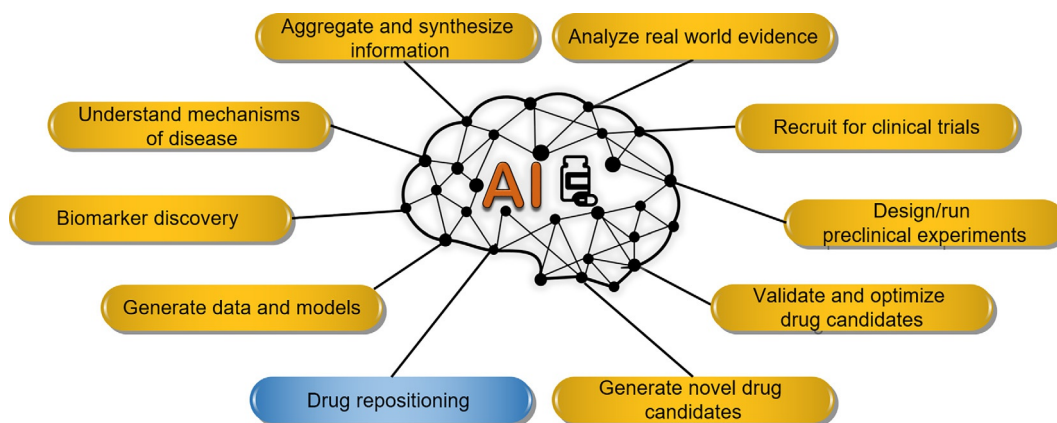

Fig. 1 AI and its application in different tasks in drug discovery and development.

Table 1   Selected examples of COVID-19-related data resources.

Data resources	Weblink	Notes
Disease etiology		
Coronavirus replicate cycle	https://science.sciencemag.org/content/367/6485/1412/tab-pdf	Efforts of current drug repositioning for COVID-19
Structure information		
Protease enzyme of COVID-19 (PDB ID: 6LU7)	https://www.rcsb.org/structure/6LU7	The crystal structure of COVID-19 main protease in complex with an inhibitor N3
SARS-CoV S-protein in complex with ACE2 (PDB ID: 2AJF)	https://www.rcsb.org/structure/2AJF	Structure of SARS coronavirus spike receptor-binding domain complexed with its receptor
Protein information		
119 Human CoV-associated host proteins	https://static-content.springer.com/esm/art%3A10.1038%2Fs41421-020-0153-3/MediaObjects/41421_2020_153_MOESM1_ESM.pdf	Zhou et al. (2020a)
332 high-confidence SARS-CoV-2-human protein-protein interactions	https://static-content.springer.com/esm/art%3A10.1038%2Fs41586-020-2286-9/MediaObjects/41586_2020_2286_MOESM7_ESM.xlsx	Gordon et al. (2020)
Pathogen genome		
Pathogen genome data of SARS-CoV-2	https://nextstrain.org/help/coronavirus/SARS-CoV-2	Nextstrain is an open-source project to harness the scientific and public health potential of pathogen genome data
Ongoing clinical trials		
Clinicaltrial.gov	https://clinicaltrials.gov/ct2/results?cond=COVID-19	Ongoing clinical trials for COVID-19
Clinical trials from WHO	https://clinicaltrials.gov/ct2/who_table	COVID-19 Studies from the World Health Organization Database
Therapeutic Options for COVID-19 Currently Under Investigation	https://covid19treatmentguidelines.nih.gov/therapeutic-options-under-investigation/	NIH: Therapeutic Options for COVID-19 Currently Under Investigation
COVID-19 ontologies		
COVID-19 ontology	https://github.com/virtual-biohackathons/covid-19-bh20/wiki/Ontology	Update ontologies with COVID-19- and CoV-related concepts to make knowledge machine-readable and actionable for mining, logical reasoning, etc. . .
Literatures		
PubMed	https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge	A collection of literatures related to COVID-19
Vulnerable population-related COVID-19		
Vulnerable populations for COVID-19	https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html	Vulnerable population for COVID-19 from CDC

Table 2   Selected data resources of Drug information.

Database	Weblink	Notes
Drug-target relationship		
DrugBank	https://www.drugbank.ca/	The latest release of DrugBank (version 5.1.4, released) contains 13,341 drug entries including 2596 approved small molecule drugs, 1289 approved biotech (protein/peptide) drugs, 130 nutraceuticals and over 6305 experimental drugs. Additionally, 5199 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries
STITCH	http://stitch.embl.de/	A resource to explore known and predicted interactions of chemicals and proteins
Therapeutic Target Database (TTD)	http://bidd.nus.edu.sg/group/cjttd/	Therapeutic Target Database (TTD) is a database to provide information about the known and explored therapeutic protein and nucleic acid targets, the targeted disease, pathway information, and the corresponding drugs directed at each of these targets
Drug transcriptomic profiles		
GEO	https://www.ncbi.nlm.nih.gov/geo/	GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are included
The Connectivity Map (CMap)	https://www.broadinstitute.org/cmap/	Provides comprehensive drug transcriptional responses of 1309 drugs or lead compounds in the clinical trials to six or seven different cancer cell lines
Open TG-GATEs	http://toxico.nibio.go.jp/english/index.html	TG-GATEs consists of the comprehensive toxicogenomic profiles of 170 compounds with four different assay types (human/rat in vitro/vivo) and multiple time and dose points in rat liver and kidney. The histopathological profiles for compounds are also available
DrugMatrix	https://ntp.niehs.nih.gov/drugmatrix/index.html	DrugMatrix contains toxicogenomic profiles for 638 different compounds from both Codelink and Affymetrix platforms, which covers multiple organisms including liver, kidney, heart, bone marrow, spleen, and skeletal muscle
LINCS – L1000	http://www.lincsproject.org/	The Library of Integrated Network-Based Cellular Signatures (LINCS) Program aims to create a network-based understanding of biology by cataloging changes in gene expression and other cellular processes that occur when cells are exposed to a variety of perturbing agents
Bioassay data		
PubChem	https://pubchem.ncbi.nlm.nih.gov/	PubChem is the world's largest collection of freely accessible chemical information. Search chemicals by name, molecular formula, structure, and other identifiers. Find chemical and physical properties, biological activities, safety and toxicity information, patents, literature citations and more
ChEMBL	https://www.ebi.ac.uk/chembl/	ChEMBL is a manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity, and genomic data to aid the translation of genomic information into effective new drugs
Drug safety data		
SIDER 4	http://sideeffects.embl.de/	SIDER contains information on marketed medicines and their recorded adverse drug reactions. The information is extracted from public documents and package inserts
FDALabel™	https://nctr-crs.fda.gov/fdalabel/ui/search	The FDALabel Database is a free web-based application used to query a database of over 100,000 drug labels
FDA Adverse Event Reporting System (FAERS)	https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard	FAERS is a computerized information database designed to support the FDA's post-marketing safety surveillance program for all approved drug and therapeutic biologic products
PharmaPendium	https://www.pharmapendium.com/login/email	PharmaPendium provides comparative regulatory document-based evidence in a single translational database for better-informed risk-benefit analyses and drug candidate assessments

Structure-based drug repurposing

Structure-based drug repurposing aims to prioritize repurposing candidates based on their affinity to disease-causing proteins (Adasme et al., 2021). Although conventional drug discovery strongly advocates the concept of "single drug-single target-single disease," it is expected that the drug may hit multiple targets. Structure-based drug repurposing looks for the off-target effect of approved drugs and investigational compounds to explore their alternative treatment possibilities. There are mainly two kinds of approaches to conducting structure-based drug repurposing, including molecular docking (Kumar and Kumar, 2019) and similarity ensemble approaches (SEA) (Keiser et al., 2007).

Molecular docking-based drug repositioning approaches comprise two underlying hypotheses: (1) if the disease-related protein target is known, then a set of small molecules or existing drugs could be docked to the known protein to prioritize the probability of the drugs to treat the disease; (2) if a small molecule is synthesized with the aim of looking for a potential pharmacological application, then reverse docking screening could be implemented to prioritize the most probable disease-related targets that bind the small molecule. High-resolution SARS-CoV-2 viral protein structures have been crystalized and stored in the RCSB Protein Databank (www.RCSB.org/covid19) (Berman et al., 2000), which is the foundation of molecular docking to screen existing molecules that may have antiviral activity against SARS-CoV-2 (Muratov et al., 2021).

SEA is an off-target prediction approach to prioritize the perturbation ability of compounds on specific proteins based on their chemical similarity among their bound ligands (Keiser et al., 2009). SEA has been widely applied to off-target identification and off-target side effects detection (Lounkine et al., 2012). Since the beginning of the COVID-19 pandemic, enormous efforts have been led by NIH/NCATS on quantitative high-throughput screening (HTS) of drugs and investigational compounds for their anti-SARS-CoV-2 activities (Chen et al., 2020a). More than 8800 compounds have been screened at four different concentrations in a SARS-CoV-2 cytopathic effect (CPE) assay using Vero E6 cells with high ACE2 (SARS-CoV-2 receptor) expression, with an accompanying cytotoxicity counter-assay and stored in the NCATS OpenData Portal (https://opendata.ncats.nih.gov/covid19/index.html) (Brimacombe et al., 2020). SARS-CoV-2 specific SEA could be developed based on these valuable data sets to look for repurposing candidates with anti-SARS-CoV-2 potency.

Genomics-based drug repurposing

The assumption of genomic-based drug repurposing is that (1) if a drug transcriptomic signature negatively correlates with a disease signature, the drug has the potential to treat the disease; (2) if the transcriptomic response signatures between the two drugs positively correlates, the indications of the two drugs could be interchangeable (Iorio et al., 2013). Advances in emerging genomics technologies have provided unprecedented pace and resolution to uncover underlying disease mechanisms at the molecular level. Multi-omics profiles such as transcriptomics, proteomics, and metabolomics have been extensively applied for causative disease gene identification, mechanistic understanding, and biomarker discovery. Notably, the Connectivity Map (CMap) and the LINCS L1000 project (Subramanian et al., 2017) are resources that have been widely applied to genomics-based drug repurposing (Qu and Rajpal, 2012; Iorio et al., 2010; Dudley et al., 2011). For example, Dudley et al. (2011) employed the genomics-based drug repurposing principle by comparing drug transcriptomic profiles with inflammatory bowel disease-associated gene signatures to enrich repurposing candidates.

Drug transcriptomic profiles have been generated in a high-throughput manner with continuing decreased costs. For example, the LINCS L1000 project has generated transcriptomic profiles of over 20 K compounds across more than 50 different cell lines. Furthermore, transcriptomic profiles have been generated across different organ systems for COVID-19 patients in different morphological stages (i.e., mild, moderate, severe) (Chen et al., 2020b). More importantly, high-resolution single-cell RNAseq data were also explored to uncover transcriptomic responses in different cell types (e.g., immune cells) of COVID-19 patients (Stephenson et al., 2021). COVID-19 is a precision medicine problem. As defined below the potential for severe COVID-19 pneumonia, extrapulmonary manifestations have also been observed and reported in patients with COVID-19 with various pre-existing conditions (Gupta et al., 2020). Furthermore, the diversity of the immune response results in the expression of various symptoms and clinical outcomes in COVID-19 patients. With these generated transcriptomic profiles associated with COVID-19 and drugs, researchers may be able to prioritize approved or investigational drugs with the most favorable profiles for treatment or prevention of COVID-19.

Network pharmacology-based drug repurposing

Network pharmacology-based drug repurposing is one of the most well-established approaches of integrating systems biology and bioinformatics to unravel the complex relationships among drugs, targets, and diseases (Lotfi Shahreza et al., 2017). In network pharmacology, the interactions (i.e., edges) among different biological concepts (i.e., nodes) can be constructed either by experimental results or statistical measures. For example, nodes represent various biological entities such as drugs, diseases, and genes. Edges represent the association between entities, such as the binding affinity of drugs on targets. Once the network is developed, the hidden relationship between drugs and diseases can be predicted using various link prediction algorithms to yield repurposing opportunities (Lotfi Shahreza et al., 2017; Badkas et al., 2021).

Furthermore, network modeling is also a fit-for-purpose drug repurposing approach for polypharmacology (Reddy and Zhang, 2013), where drug combination treatments are explored to enhance treatment efficacy. For example, complex viral diseases such as

HIV-1 and HCV involve different biological processes and interfere with various targets and pathways. A single drug may not be sufficient to limit the emergence of drug resistance during chronic viral replication, and drug combinations of non-overlapping targets may provide more effective antiviral treatment (Zhang et al., 2016). In the past two decades, tremendous progress has been made in bioengineering, resulting in high-resolution data profiles representing biological complexes related to disease etiology and pathogenesis. These data profiles such as protein-protein interactions (PPIs) (Szklarczyk et al., 2021), drug-target relationships (DTIs) (Wishart et al., 2018; Mendez et al., 2019), disease and disease associations (Köhler et al., 2019) have been utilized for network modeling in developing treatments. Encouragingly, 332 high-quality SARS-CoV-2 regulating host proteins have been identified using affinity-purification–mass spectrometry (AP-MS)-based proteomics (Gordon et al., 2020). Furthermore, some host proteins associated with human coronaviruses (HCoVs), including SARS-CoV, MERS-CoV, IBV, MHV, HCoV-229E, and HCoV-NL63 were also curated (Zhou et al., 2020a). These data may be utilized to develop SARS-CoV-2 specific network modeling to facilitate the COVID-19 repurposing candidate identification.

Mechanism driven drug repurposing

Mechanism-driven drug repurposing seeks repurposing candidates based on specific underlying mechanism or hypothesis of the disease pathogenesis or evolution. For example, side effects are clinical outcomes of patient phenotypic responses when taking certain drugs, which could help discover new therapeutic uses (Campillos et al., 2008). The rationale behind side effect-based drug repurposing is that if two drugs share similar side effect profiles, they may share the same off-target and therapeutic uses (Yang and Agarwal, 2011; Bisgin et al., 2014). The phenome-wide association study (PheWAS) is also an essential mechanism-driven drug repurposing strategy, which aims to explore new genetic variants and disease associations based on clinical evidence embedded in a large number of electronic medical records (EMRs). Subsequently, the potential treatment options of the disease may be identified if a drug could perturb disease-associated genetic variants (Denny et al., 2013).

Global efforts to conduct different mechanistic studies have provided insight into the pathogenesis of COVID-19. Consequently, hundreds of thousands of scientific publications are publicly available, and clinical trials data are stored in clinical repositories, which are rich resources for developing potential COVID-19 treatments. For example, the CORD-19 dataset has been curated to create a machine-readable COVID-19 literature collection. Additionally, as of October 19, 2021, over 6800 COVID-19 clinical trials are listed on the clinicaltrials.gov website, providing details such as study design, interventions, trial locations, and phase status (https://clinicaltrials.gov/ct2/results?cond=COVID-19). These resources can help researchers generate different repurposing hypotheses by taking full advantage of AI-powered language models to mine valuable information to facilitate AI-powered drug repurposing development (Liu et al., 2021a).

The opportunity of AI-powered drug repurposing against COVID-19

AI makes it possible for machines to learn from experience, adjust to new inputs and perform human-like tasks. Different types of deep learning algorithms are now available which provide a unique angle to extract and make sense of useful information behind the data (Fig. 2). First, deep learning can integrate information from heterogeneous data sources with a newly generated high-level
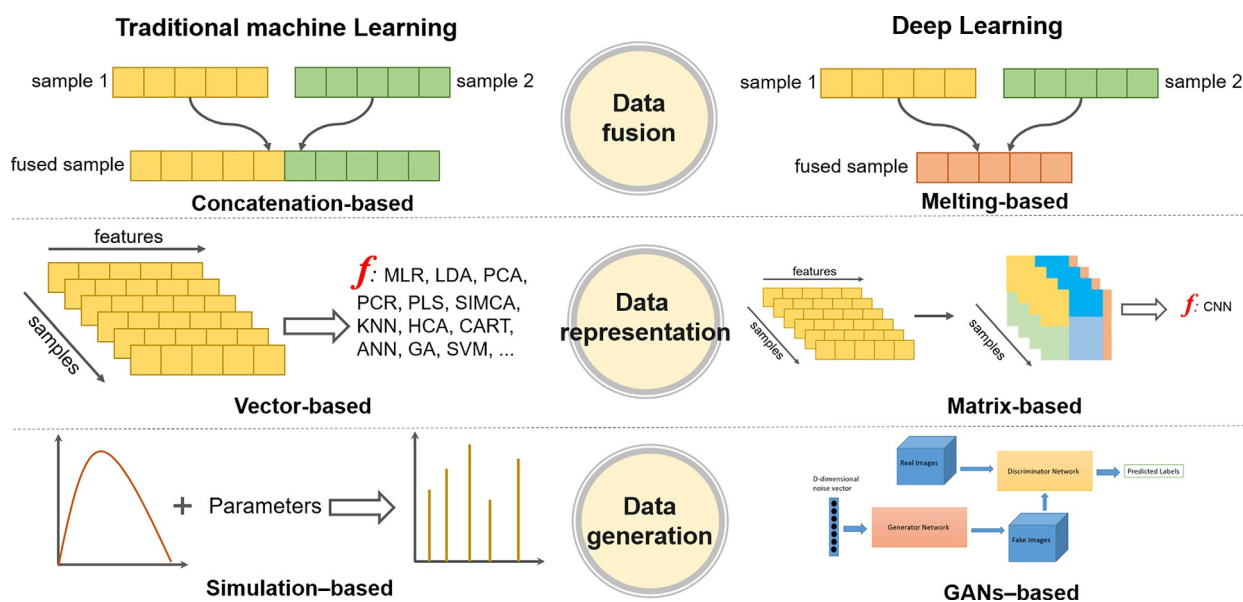

Fig. 2 A comparison between machine learning and deep learning.

representation, enabling a new data fusion method. Second, deep learning provides a novel way to organize data for model development. Unlike the engineered feature-based data representation widely applied in machine learning algorithms, deep learning can directly extract information from matrix-based data representations by considering the relationship between coordinate neighbors. Lastly, deep learning approaches, such as generative adversarial networks (GANs), innovates the generative model structure from statistical data simulation to a generator and discriminator competition process. The merits of different deep learning frameworks are increasingly combined with other drug repurposing principles to provide more robust and efficient therapy development solutions. Here, we highlight potential opportunities for AI-powered drug repurposing in combating COVID-19.

Opportunity 1: AI accelerates and enhances structure-based drug repurposing

A high-resolution protein structure is the basis of a reliable molecular docking protocol. The genome of SARS-COV-2 encodes many essential proteins, including the nucleocapsid (N) protein, Spike (S) protein, Envelope (E) protein, Membrane (M) protein, and coronavirus main protease for its replication in the host genome, which plays an essential role in cleaving polyproteins into replication-related proteins and regulation of gene expression (Prasad and Prasad, 2020). Among these encoded viral proteins, heavily glycosylated S trimers bind to the angiotensin-converting enzyme 2 (ACE-2) receptor and mediate entry of virions into target cells (Walls et al., 2020; Shang et al., 2020; Wang et al., 2020).

The high-resolution viral structure greatly relies on advanced crystallization technology, cryo-electron microscopy, and tomography (Ke et al., 2020), which is time-consuming and costly. AI paves an innovative way to predict the structure of essential proteins crucial for virus entry and replication in a timely manner. For example, Google DeepMind developed an AlphaFold algorithm using deep residual networks (DRN) called ResNets and has successfully predicted protein structures of membrane protein, protein 3a, nsp2, nsp4, nsp6, and papain-like C-terminal domain of SARS-CoV-2, offering a huge impetus for structure-based drug repurposing (Senior et al., 2020). Furthermore, AI could help in isolating the complex structures from images of cryo-electron microscopy. A customized deep convolutional neural network framework titled DeepTracer was developed and applied to derive the SARS-CoV-2 protein structure from high-resolution cryo-electron microscopy density maps (Pfab et al., 2021).

Although molecular docking remains a popular method for the virtual screening of ligands for exploring their potential therapeutic uses, its power is limited owing to the high computational cost and vast chemical space. Machine learning, particularly deep learning, has shown great promise in aiding molecular docking-based drug repurposing by modeling docking scores with chemical information (Mohapatra et al., 2020; Srinivasan et al., 2021; Hu et al., 2020). Srinivasan et al. proposed a Monte Carlo tree search algorithm combined with a multitask neural network surrogate model to predict Vina docking scores that prioritize the binding affinity of ligands to the S-protein of the SARS-CoV-2 virus (Srinivasan et al., 2021). The deep learning models for predicting docking scores only require a few hundred or thousands of drugs as a training set. Then the trained models could be utilized to estimate docking scores for millions of compounds covering vast chemical space without an actual docking process. However, considering the developed deep learning model based on limited drugs, further investigation on the generalization of this approaches is needed. Because the generalizability or robustness of deep learning to unknown data or data coming from a different distribution is an ongoing issue. Furthermore, Deep learning bias is also a problem in many DL applications, especially when training data is limited.

SEA identifies "progeny" compounds (i.e., repurposing candidates) that are similar to the "parent" compounds already being tested against specific targets. The similarity between "progeny" and "parent" compounds are typically based on chemical-physical proprieties-based descriptors (e.g., Mold2 (Hong et al., 2008) and Dragon (Mauri et al., 2006)), and topological fingerprints (e.g., extended connectivity fingerprints – ECFPs (Rogers and Hahn, 2010)). Inspired by considerable success in AI-powered natural language processing (NLP), AI-based chemical representation is proposed for various downstream tasks such as predictive model development and chemical similarity assessment. Specifically, the simplified molecular-input line-entry system (SMILES), short ASCII strings, are considered as text-based notation for describing the structure of chemical species. Then, different AI algorithms such as autoencoders (Jaeger et al., 2018) and Bidirectional Encoder Representations from Transformers (BERT) (Wang et al., 2019) could be applied to learn the SMILES and then project them into the latent space to represent chemical information. The generated dense vector representations are conservative values, overcoming drawbacks of conventional chemical descriptors suffering from sparseness and bit collisions. Moskal et al. (2020) conducted a comprehensive assessment of AI-based chemical representation, including the linguistics-inspired Mol2Vec and the Estimated Shape Representation (ESR), for their performance on finding drugs to treat COVID-19.

Opportunity 2: AI enables reversed engineering to generate lead compounds based on gene expression

Genomics-based drug repurposing, also called the CMap approach, is widely adopted since the transcriptomic profiles can characterize cellular and organismal phenotypes of drugs and diseases (Lussier and Chen, 2011; Iwata et al., 2017; Ge et al., 2021). High-throughput screening (HTS) technologies have tremendously improved the data generation speed, resulting in a significant expanse of the compounds and cell types from the CMap project (e.g., ~1300 compounds and ~6 cancer cell lines) to

the LINCS 1000 project (e.g., ~20,000 compounds and ~50 cell lines) (Subramanian et al., 2017). However, there are still a few gaps for fully implementing comprehensive drug repurposing assessment with the transcriptomic profiles in the current version of the LINCS project: (1) more compounds were screened in some cell lines than others; and (2) there are still a large number of high drug-like, purchasable chemicals not included.

AI could be an alternative means to computationally infer the biological profiles to facilitate genomics-based drug repurposing, including two scenarios: (1) infer transcriptomic profiles at specific experimental conditions (i.e., duration time/dose/cell culture) based on chemical information; and (2) de novo drug design based on drug transcriptomic profiles. Pham et al. (2021) proposed a DeepCE method by integrating a graph convolutional network (GCN) and a multilayer feed-forward neural network to predict the differential gene expression profile perturbed by de novo chemicals in the LINCS project and applied it to COVID-19 drug repurposing. The proposed DeepCE generated a few repurposing candidates (e.g., faldaprevir and alisporivir) that are in ongoing clinical trials for COVID-19 efficacy. Méndez-Lucio et al. (2020) reported a generative adversarial network (GAN) model to de novo generate active-like molecules based on desired transcriptomic profiles. The proposed GAN model could be used in drug repurposing to generate disease-specific lead compounds based on disease signatures of COVID-19 patients.

Opportunity 3: AI enhances network pharmacology-based drug repurposing

Conventional network pharmacology-based drug repurposing deploys hidden link prediction (i.e., novel drug and disease associations) using a statistical measure of topological information from the known association among biological entities (Badkas et al., 2021; Stolfi et al., 2020; Liu et al., 2021b; Li et al., 2021; Habibi and Taheri, 2021). Zhou et al. (2020a) adopted an integrative network pharmacology-based drug repurposing framework to look for potential single and combinational drugs for COVID-19. A list of 16 repurposing candidates for possible anti-HCoV activity was enriched and further verified with transcriptomic analysis in human cell lines. The proposed approach offered a rapid technique for repurposing candidate identification based on network proximity analysis topological relationship between drugs and targets.

A biological network could also be considered a knowledge graph. Therefore, graph-based deep learning strategies (Geyer, 2017) could be employed to extract the hidden relationship among biological entities. Morselli Gysi et al. (2021) developed hybrid network pharmacology approaches by integrating graph convolutional networks (GCNs), network diffusion, and network proximity based on consensus PPI and drug-target relationship data to prioritize 6340 drugs regarding their efficacy against SARS-CoV-2.

Some opportunities are worth investigating for further improving network-network pharmacology-based drug repurposing with AI. For example, the PPI network could be constructed based on co-expression, gene fusion, co-occurrences, etc. Different association types could be weighted in the network modeling (e.g., Graph Neural Networks GNNs), which may provide more resolutions for hidden relationship extractions for COVID-19 drug repurposing. A set of GNNs such as DeepWalk, Node2vec, LINE, and GraphSAGE gains momentum in network analysis. These GNNs have a solid ability to represent the node information into a latent vector space by capturing the neighborhood properties in the network, which is worth further investigation for enhancing network pharmacology-based drug repurposing.

AI could help integrate diverse heterogeneous biological networks for enhanced drug repurposing. Zeng et al. (2020) developed an AI-driven drug repurposing method known as deepDTnet that utilizes heterogeneous biological network data among different biological entities to predict new interactions between drugs and targets with greater accuracy than previous methods. This methodology embeds a network connecting drugs, targets, and diseases via deep learning (e.g., autoencoders) to infer new drug and target association, which may also be helpful in COVID-19 repurposing.

Opportunity 4: AI facilitates the mechanism-based drug repurposing

Accumulated knowledge on the natural history of SARS-CoV-2 is an important resource for developing treatments. Biomedical literature is a primary source underpinning diverse information regarding COVID-19 and embedded repurposing opportunities. BenevolentAI developed an extensive knowledge graph consisting of a large repository of structured medical information and their interior relationship by using Monte Carlo tree search and symbolic artificial intelligence (AI) approaches (Segler et al., 2018). BenevolentAI identified baricitinib, originally approved to treat rheumatoid arthritis, as a potential therapy for COVID-19, which might block the viral infection process and reduce the ability of the virus to infect lung cells (Richardson et al., 2020). Furthermore, there were 2 randomized, double-blind, placebo-controlled clinical trials that supported the efficacy and demonstrated a mortality benefit (https://www.fda.gov/media/143823/download). Encouragingly, the US FDA has granted emergency use authorization (EUA) to baricitinib for the treatment of COVID-19 in those hospitalized adults and pediatric patients 2 years of age or older requiring supplemental oxygen, non-invasive or invasive mechanical ventilation, or extracorporeal membrane oxygenation (ECMO) (https://www.fda.gov/media/143822/download).

NLP, as a critical drug repurposing strategy, aims to extract the hidden relationships between drugs and diseases from free text-based biomedical documents (Zhang et al., 2021). AI-powered language models (LMs) have changed the landscape of NLP fields. Notably, different transformer based LMs such as BERT and its derivatives outperformed state-of-the-art NLP approaches in

various NLP tasks, including text classification, named entity recognition (NER), question and answering (Q&A), and text summarization (Liu et al., 2021a). Zhang et al. (2021) proposed literature-derived knowledge and knowledge-graph completion methods using BERT and five different neural knowledge-graph completion algorithms for COVID-19 repurposing. They found five highly ranked drugs regarding potential COVID-19 treatment were mechanically well explained.

AI-powered LMs can mimic human intelligence to learn the knowledge and apply it to solve real-world questions. Notably, domain-specific training is an effective way to improve information retrieval. For example, BioBERT was developed based on the BERT model with further training using biomedical literature, which outperformed the original BERT in various biomedical NLP tasks. A similar strategy was also adopted to develop ClinicalBERT with EMRs. Müller et al. (2020) developed COVID-Twitter-BERT (CT-BERT) based on a large corpus of Twitter messages on the topic of COVID-19, improving 10–30% over BERT-base and large models in text classification task in five independent datasets. Further investigation on CT-BERT and other COVID-19 specific LMs may hold promise in improving the performance of mechanism-based drug repurposing.

Outlook

AI-driven drug repurposing has been shown during the COVID-19 pandemic to accelerate therapy development and push new computing paradigms' boundaries to real-world applications (Zhou et al., 2020b). However, concerns have been raised on COVID-19 data generated from inadequate peer review, which could mislead drug development efforts, confuse policymakers, and compromise clinical practice (Levin et al., 2020). Therefore, the data employed in AI model development should be adequately scrutinized. Some large consortium- and federal government-led efforts such as Open-Access Data and Computational Resources to Address COVID-19 (https://datascience.nih.gov/covid-19-open-access-resources) aim to provide a centralized COVID-19 data atlas, which may improve data quality.

Furthermore, robust data sharing principles are highly recommended to enable machines to automatically find and use COVID-19 data in a secured environment. FAIR provided a standardized data management infrastructure to ensure public biomedical data meet required principles of findability, accessibility, interoperability, and reusability (Wilkinson et al., 2016). Some initial attempts led by the VODAN core consortium aim to design and rapidly build a genuinely international and interoperable, distributed data FAIR-based network infrastructure that supports evidence-based responses to the COVID-19 pandemic (Mons, 2020). More efforts to develop standardized data management frameworks are urgently needed to warrant establishing a high-quality AI-powered drug repurposing strategy.

Although AI-driven drug repurposing may provide a list of prioritized repurposing candidates, it is a great challenge to experimentally verify the efficacy and safety in the clinical setting due to a lack of reliable data (Delavan et al., 2018). Typically, repurposing candidates from computational approaches were affirmed based on in vitro and in vivo models, literature surveys, or ongoing clinical trials. Very few candidates were directly adopted to be confirmed with a newly designed clinical trial. Thus, there is a lack of standardized measures to evaluate the performance of AI-driven drug repurposing approaches. Also, it is not fair to compare drug repurposing approaches with different principles since they employ different data profiles and algorithms. Therefore, it is imperative to combine AI-based drug repurposing approaches with other scientific disciplines to improve clinical adoption. Meanwhile, as many current algorithms were mainly focusing on the affinity between the drug and the target, it will be important to also consider other information such as the safely achievable drug concentration in vivo (Fan et al., 2020).

Although AI has made significant progress in innovating biomedical fields, the gap still exists to leverage the success of AI in the regulatory setting. Multi-government agencies are promoting and advocating to develop robust, safe, secure, and privacy-preserving machine learning to prioritize fundamental and translational AI research consistent with the Administration's priorities (https://www.whitehouse.gov/wp-content/uploads/2021/07/M-21-32-Multi-Agency-Research-and-Development-Priorities-for-FY-2023-Budget-.pdf). It is an open question on how AI-driven drug repurposing might be positioned in the proper context of regulatory application. We hope this chapter will trigger different stakeholders' interest in further standardizing, enhancing, and promoting AI-powered drug repurposing and providing extra value in drug development for emerging infections.

Acknowledgments

References

Adasme MF, et al. (2021) Structure-based drug repositioning: Potential and limits. *Seminars in Cancer Biology* 68: 192–198.
Arora N, et al. (2020) The role of artificial intelligence in tackling COVID-19. *Future Virology*. https://doi.org/10.2217/fvl-2020-0130.
Badkas A, et al. (2021) Topological network measures for drug repositioning. *Briefings in Bioinformatics* 22(4).
Bavel JJV, et al. (2020) Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour* 4(5): 460–471.

Berman HM, et al. (2000) The protein data bank. *Nucleic Acids Research* 28(1): 235–242.

Bisgin H, et al. (2014) A phenome-guided drug repositioning through a latent variable model. *BMC Bioinformatics* 15(1): 267.

Brimacombe KR, et al. (2020) An OpenData portal to share COVID-19 drug repurposing data in real time. *BioRxiv.* https://doi.org/10.1101/2020.06.04.135046.

Campillos M, et al. (2008) Drug target identification using side-effect similarity. *Science* 321(5886): 263–266.

Chen CZ, Shinn P, Itkin Z, Eastman RT, Bostwick R, Rasmussen L, Huang R, Shen M, Hu X, Wilson KM, Brooks B, Guo H, Zhao T, Klump-Thomas C, Simeonov A, Michael SG, Lo DC, Hall MD, and Zheng W (2020a) Drug repurposing screen for compounds inhibiting the cytopathic effect of SARS-CoV-2. *bioRxiv: The Preprint Server for Biology.* https://doi.org/10.1101/2020.08.18.255877. 2020.08.18.255877.

Chen Y-M, et al. (2020b) Blood molecular markers associated with COVID-19 immunopathology and multi-organ damage. *The EMBO Journal* 39(24): e105896.

Cutler DM and Summers LH (2020) The COVID-19 pandemic and the $16 trillion virus. *JAMA* 324(15): 1495–1496.

Delavan B, et al. (2018) Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discovery Today* 23(2): 382–394.

Denny JC, et al. (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature Biotechnology* 31(12): 1102–1110.

Dudley JT, et al. (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Science Translational Medicine* 3(96): 96ra76.

Fan J, et al. (2020) Connecting hydroxychloroquine in vitro antiviral activity to in vivo concentration for prediction of antiviral effect: A critical step in treating patients with coronavirus disease 2019. *Clinical Infectious Diseases* 71(12): 3232–3236.

Fang H, et al. (2020) FDALabel for drug repurposing studies and beyond. *Nature Biotechnology* 38(12): 1378–1379.

Ge Y, et al. (2021) An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *Signal Transduction and Targeted Therapy* 6(1): 165. 2020.2003.2011.986836.

Geyer F (2017) Performance evaluation of network topologies using graph-based deep learning. In: *Proceedings of the 11th EAI International Conference on Performance Evaluation Methodologies and Tools*, pp. 20–27. Association for Computing Machinery.

Gil C and Martinez A (2021) Is drug repurposing really the future of drug discovery or is new innovation truly the way forward? *Expert Opinion on Drug Discovery* 16(8): 829–831.

Gordon DE, et al. (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583(7816): 459–468.

Gupta A, et al. (2020) Extrapulmonary manifestations of COVID-19. *Nature Medicine* 26(7): 1017–1032.

Habibi M and Taheri G (2021) Topological network based drug repurposing for coronavirus 2019. *PLoS One* 16(7): e0255270.

Harrison C (2020) Coronavirus puts drug repurposing on the fast track. *Nature Biotechnology* 38(4): 379–381.

Harvey WT, et al. (2021) SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology* 19(7): 409–424.

Hong H, et al. (2008) Mold2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *Journal of Chemical Information and Modeling* 48(7): 1337–1344.

Hu F, et al. (2020) *Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model. arXiv preprint arXiv:2003.00728.*

Iorio F, et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proceedings of the National Academy of Sciences* 107(33): 14621–14626.

Iorio F, et al. (2013) Transcriptional data: A new gateway to drug repositioning? *Drug Discovery Today* 18(7–8): 350–357.

Iwata M, et al. (2017) Elucidating the modes of action for bioactive compounds in a cell-specific manner by large-scale chemically-induced transcriptomics. *Scientific Reports* 7(1): 40164.

Jaeger S, et al. (2018) Mol2vec: Unsupervised machine learning approach with chemical intuition. *Journal of Chemical Information and Modeling* 58(1): 27–35.

Ke Z, et al. (2020) Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* 588(7838): 498–502.

Keiser MJ, et al. (2007) Relating protein pharmacology by ligand chemistry. *Nature Biotechnology* 25(2): 197–206.

Keiser MJ, et al. (2009) Predicting new molecular targets for known drugs. *Nature* 462(7270): 175–181.

Klug DM, et al. (2016) Repurposing strategies for tropical disease drug discovery. *Bioorganic & Medicinal Chemistry Letters* 26(11): 2569–2576.

Köhler S, et al. (2019) Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research* 47(D1): D1018–D1027.

Kumar S and Kumar S (2019) Chapter 6 - Molecular docking: A structure-based approach for drug repurposing. In: Roy K (ed.) *In Silico Drug Design*, pp. 161–189. Academic Press.

Levin JM, et al. (2020) Artificial intelligence, drug repurposing and peer review. *Nature Biotechnology* 38(10): 1127–1131.

Li L, et al. (2020) Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy. *Radiology* 296(2): E65–E71. https://doi.org/10.1148/radiol.2020200905.

Li X, et al. (2021) Network bioinformatics analysis provides insight into drug repurposing for COVID-19. *Medicine in Drug Discovery* 10: 100090.

Liu Z, et al. (2013) In silico drug repositioning – what we need to know. *Drug Discovery Today* 18(3): 110–115.

Liu Z, et al. (2021a) AI-based language models powering drug discovery and development. *Drug Discovery Today* 26(11): 2593–2607.

Liu DY, et al. (2021b) Drug repurposing for COVID-19 treatment by integrating network pharmacology and transcriptomics. *Pharmaceutics* 13(4).

Lotfi Shahreza M, et al. (2017) A review of network-based approaches to drug repositioning. *Briefings in Bioinformatics* 19(5): 878–892.

Lounkine E, et al. (2012) Large-scale prediction and testing of drug activity on side-effect targets. *Nature* 486(7403): 361–367.

Lussier YA and Chen JL (2011) The emergence of genome-based drug repositioning. *Science Translational Medicine* 3(96). 96ps35.

Maghdid H, et al. (2020) A novel AI-enabled framework to diagnose coronavirus COVID 19 using smartphone embedded sensors. *Design Study* 1–7.

Mauri A, et al. (2006) Dragon software: An easy approach to molecular descriptor calculations. *Match* 56(2): 237–248.

Mendez D, et al. (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research* 47(D1): D930–d940.

Méndez-Lucio O, et al. (2020) De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nature Communications* 11(1): 10.

Miner AS, et al. (2020) Chatbots in the fight against the COVID-19 pandemic. *NPJ Digital Medicine* 3: 65.

Mohapatra S, et al. (2020) Repurposing therapeutics for COVID-19: Rapid prediction of commercially available drugs through machine learning and docking. *PLoS One* 15(11): e0241543.

Mons B (2020) The VODAN IN: Support of a FAIR-based infrastructure for COVID-19. *European Journal of Human Genetics* 28(6): 724–727.

Morselli Gysi D, et al. (2021) Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proceedings of the National Academy of Sciences* 118(19): e2025581118.

Moskal M, et al. (2020) *Suggestions for second-Pass Anti-COVID-19 Drugs Based on the Artificial Intelligence Measures of Molecular Similarity, Shape and Pharmacophore Distribution*. Cambridge: Cambridge Open Engage, ChemRxiv.

Müller M, et al. (2020) COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on twitter. *ArXiv.* abs/2005.07503.

Muratov EN, et al. (2021) A critical overview of computational approaches employed for COVID-19 drug discovery. *Chemical Society Reviews* 50(16): 9121–9151.

Panch T, et al. (2019) Artificial intelligence: opportunities and risks for public health. *The Lancet Digital Health* 1(1): e13–e14.

Parasher A (2021) COVID research: A year of scientific milestones. *Nature.* https://www.nature.com/articles/d41586-020-00502-w.

Pfab J, et al. (2021) DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proceedings of the National Academy of Sciences* 118(2).

Pham T-H, et al. (2021) A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nature Machine Intelligence* 3(3): 247–257.

Planas D, et al. (2021) Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 596(7871): 276–280.

Prasad A and Prasad M (2020) SARS-CoV-2: The emergence of a viral pathogen causing havoc on human existence. *Journal of Genetics* 99(1): 1–4.

Pushpakom S, et al. (2019) Drug repurposing: Progress, challenges and recommendations. *Nature Reviews Drug Discovery* 18(1): 41–58.

Qu XA and Rajpal DK (2012) Applications of Connectivity Map in drug discovery and development. *Drug Discovery Today* 17(23–24): 1289–1298.

Reddy AS and Zhang S (2013) Polypharmacology: Drug discovery for the future. *Expert Review of Clinical Pharmacology* 6(1): 41–47.

Richardson P, et al. (2020) Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *The Lancet* 395(10223): e30–e31.

Rogers D and Hahn M (2010) Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling* 50(5): 742–754.

Segler MHS, et al. (2018) Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 555(7698): 604–610.

Senior AW, et al. (2020) Improved protein structure prediction using potentials from deep learning. *Nature* 577(7792): 706–710.

Shang J, et al. (2020) Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581(7807): 221–224.

Srinivasan S, et al. (2021) Artificial intelligence-guided de novo molecular design targeting COVID-19. *ACS Omega* 6(19): 12557–12566.

Stephenson E, et al. (2021) Single-cell multi-omics analysis of the immune response in COVID-19. *Nature Medicine* 27(5): 904–916.

Stolfi P, et al. (2020) Designing a network proximity-based drug repurosing strategy for COVID-19. *Frontiers in Cell and Development Biology* 8(1021).

Subramanian A, et al. (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171(6): 1437–1452.e1417.

Szklarczyk D, et al. (2021) The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* 49(D1): D605–D612.

Vamathevan J, et al. (2019) Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery* 18(6): 463–477.

Walls AC, et al. (2020) Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181(2): 281–292.e286.

Wang S, et al. (2019) SMILES-BERT: Large scale unsupervised pre-training for molecular property prediction. In: Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics 429–436.

Wang Q, et al. (2020) Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 181(4): 894–904.e899.

Wilkinson MD, et al. (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(1), 160018.

Wishart DS, et al. (2018) DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research* 46(D1): D1074–D1082.

Yang L and Agarwal P (2011) Systematic drug repositioning based on clinical side-effects. *PLoS One* 6(12): e28025.

Zeng X, et al. (2020) Target identification among known drugs by deep learning from heterogeneous networks. *Chemical Science* 11(7): 1775–1797.

Zhang W, et al. (2016) Polypharmacology in drug discovery: A review from systems pharmacology perspective. *Current Pharmaceutical Design* 22(21): 3171–3181.

Zhang Z, et al. (2020) Overcoming cancer therapeutic bottleneck by drug repurposing. *Signal Transduction and Targeted Therapy* 5(1): 113.

Zhang R, et al. (2021) Drug repurposing for COVID-19 via knowledge graph completion. *Journal of Biomedical Informatics* 115: 103696. https://doi.org/10.1016/j.jbi.2021.103696.

Zhou Y, et al. (2020a) Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery* 6: 14.

Zhou Y, et al. (2020b) Artificial intelligence in COVID-19 drug repurposing. *The Lancet Digital Health* 2(12): e667–e676.