

# COVID19 Drug Repository: text-mining the literature in search of putative COVID19 therapeutics

Dmitry Tworowski<sup>†</sup>, Alessandro Gorohovski<sup>†</sup>, Sumit Mukherjee, Gon Carmi, Eliad Levy, Rajesh Detroja, Sunanda Biswas Mukherjee and Milana Frenkel-Morgenstern<sup>\*</sup>

Laboratory of Cancer Genomics and Biocomputing of Complex Diseases, Azrieli Faculty of Medicine, Bar-Ilan University, Henrietta Szold 8, Safed 13195, Israel

Received August 25, 2020; Revised October 07, 2020; Editorial Decision October 08, 2020; Accepted November 04, 2020

## ABSTRACT

The recent outbreak of COVID-19 has generated an enormous amount of Big Data. To date, the COVID-19 Open Research Dataset (CORD-19), lists ~130,000 articles from the WHO COVID-19 database, PubMed Central, medRxiv, and bioRxiv, as collected by Semantic Scholar. According to LitCovid (11 August 2020), ~40,300 COVID19-related articles are currently listed in PubMed. It has been shown in clinical settings that the analysis of past research results and the mining of available data can provide novel opportunities for the successful application of currently approved therapeutics and their combinations for the treatment of conditions caused by a novel SARS-CoV-2 infection. As such, effective responses to the pandemic require the development of efficient applications, methods and algorithms for data navigation, text-mining, clustering, classification, analysis, and reasoning. Thus, our COVID19 Drug Repository represents a modular platform for drug data navigation and analysis, with an emphasis on COVID-19-related information currently being reported. The COVID19 Drug Repository enables users to focus on different levels of complexity, starting from general information about (FDA-) approved drugs, PubMed references, clinical trials, recipes as well as the descriptions of molecular mechanisms of drugs' action. Our COVID19 drug repository provide a most updated world-wide collection of drugs that has been repurposed for COVID19 treatments around the world.

## INTRODUCTION

The COVID-19 pandemic outbreak has triggered immediate reactions from the medical and scientific communities, and has resulted in an explosive growth of novel data regarding possible therapies or therapeutic oppor-

tunities (1,2). The COVID-19 data portal (<https://www.covid19dataportal.org/>) established by the European Commission in April, 2020 has facilitated the exchange and sharing of COVID-19 research data. One of the first open initiatives realized with creation of this portal was the development of the COVID-19 Open Research Dataset (CORD-19) (2). The CORD-19 (<https://www.semanticscholar.org/cord19>) currently lists ~130,000 articles from the WHO COVID-19 database, PubMed Central, medRxiv, and bioRxiv, as collected by Semantic Scholar. Another comprehensive list of COVID-19 databases and journals can be found on the Centres for Disease Control (CDC) library webpage:

<https://www.cdc.gov/library/researchguides/2019novelcoronavirus/databasesjournals.html>.

According to recent records from LitCovid resource (1), 40 300 COVID19-related articles have been currently listed in PubMed (1). The rapid accumulation of COVID-19 literature requires novel tools for the data collection and organization with efficient navigation capabilities. Such navigation capabilities are based on the literature-based discovery (LBD) concept (3) and can be achieved by implementing text-mining, clustering, and classification methods (1,4–8). Available text and data-mining tools, such as those found at LitCovid (1), PubTator (4,9,10), the iSearch platform (<https://icite.od.nih.gov/covid19/search/>), NeuralCovidex (<https://covidex.ai/>) (7), the COVID-19 Data Portal (<https://www.covid19dataportal.org/>), Carrot/Lingo (<https://search.carrot2.org/#/web>) (11) and ProtFus (12), efficiently extract target information across articles and other text sources. Using the mentioned tools for text-mining, we have created the COVID-19 Drug Repository.

The goal of our COVID-19 Drug Repository was to automatically collect data on drugs used against COVID-19 around the world and build a structured repository that includes drug descriptions, side effects and available publications. The repository also contains medicine- and pharmacology-oriented data, including annotated information on (FDA-)approved drugs, therapeutic agents (experimental drugs), and drug-like synthetic or natural chemi-

<sup>\*</sup>To whom correspondence should be addressed. Tel: +972 722644901; Email: milana.morgenstern@biu.ac.il

<sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

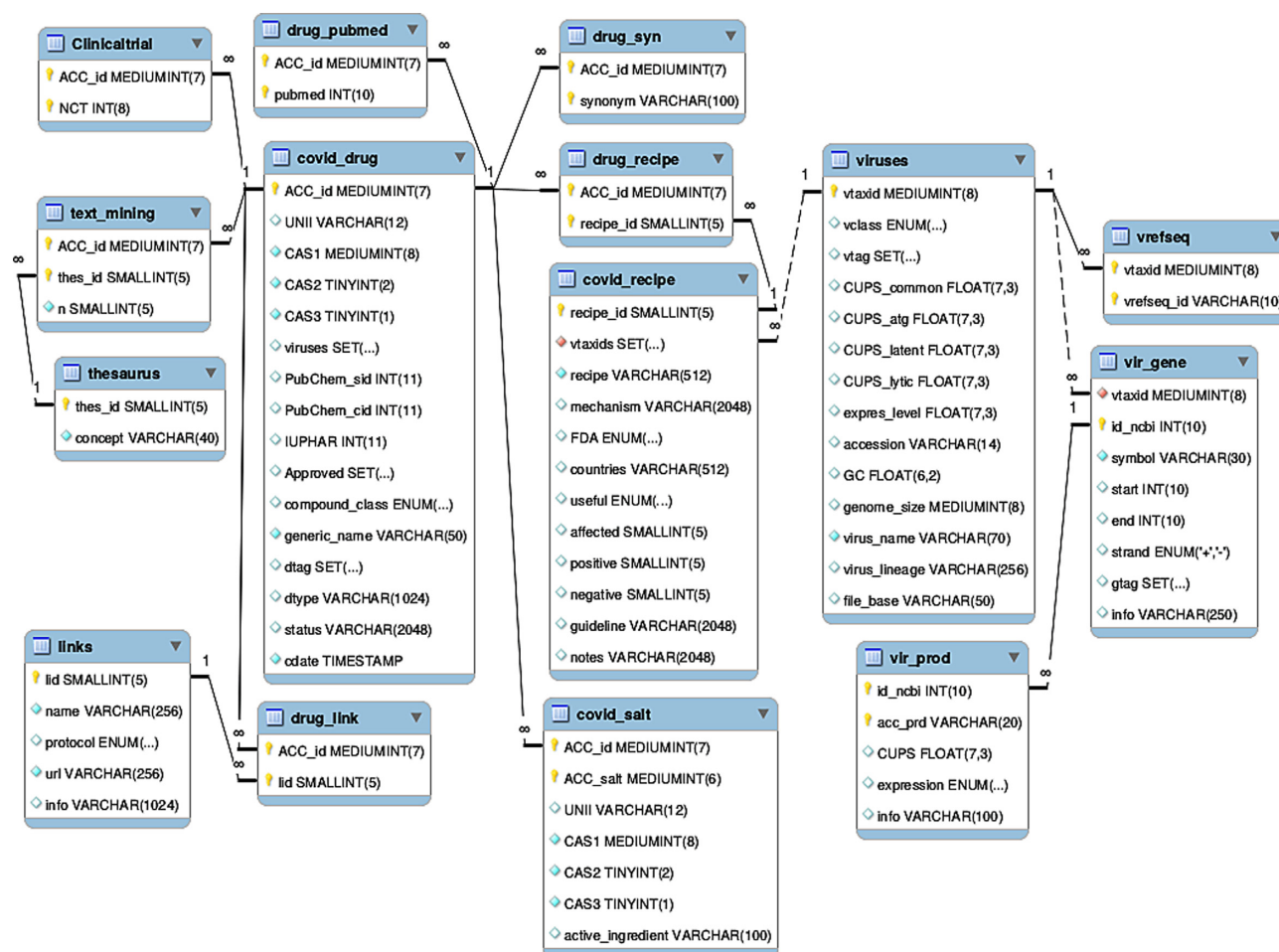


Figure 1. The structure/map of the COVID-19 Drugs Repository.

cal substances. The data was collected and integrated by methods developed for the ‘omics’ field (13), in particular, chemogenomics (i.e. chemical genomics) (14–17), pharmacogenomics (18–27), genomics and personomics (28–30). In addition, we made use of a number of chemogenomics (31–33) and pharmacogenomics (34–36) approaches that focused on the repositioning (i.e., repurposing) of FDA approved drugs and clinical trials in the treatment of COVID-19. All the data collected in the COVID-19 Drug Repository are designed for use by researchers and clinicians in the field. The information cannot be used for self-medication!

## RESULTS

### The COVID-19 Drug Repository: structure and technical description

COVID-19 Drug Repository is an open-source modular platform built on the MySQL server platform, comprising 15 curated tables. The structure of the database, presenting the logical relations between these tables, and the data collection process are shown in Figures 1 and 2, respectively. To ensure consistency between the *drug\_syn*, *drug\_recipe*, *covid\_salt*, *drug\_link*, *drug\_pubmed*, *Clinicaltrial*, *text\_mining*

and *covid\_drug* tables, the insertion/update/deletion of rows is linked to the *covid\_drug* table. Each *covid\_drug* entry is linked to 15 data fields corresponding to drug data and a target. Most of the data fields (i.e. ACC\_id, UNII, CAS1, CAS2, CAS3, and PubChem\_cid) are hyperlinked to other databases (i.e. DrugBank (37,38), ClinicalTrials.gov, PubChem (39), IUPHAR/BPC (40) and Chemical Abstracts Service (41,42) (Figure 1, Figure 2, Table 1). Each *covid\_recipe* entry is associated with 12 data fields, including drug formulation (recipe), citing the country of manufacture, FDA-approved drugs, guidelines, etc. The COVID-19 Drug Repository supports text query inputs using the search box on the homepage (Figure 3). The MyISAM engine (43) was implemented to support the FULLTEXT search functionality, with the ‘utf8’ DEFAULT CHARSET. Detailed instructions on the browsing and search tools found in the database were provided below and can also be found on the database homepage (under ‘Help’ option). Finally, the database update process is semi-automatic as follows: (a) selection of potential COVID19 therapeutic substances found in research articles is manual; (b) updating and adding new records to COVID19 Drug Repository database is fully automated using Perl scripts; (c) hyperlinks to PubMed and other sources, and maps are generated automatically by python scripts.

## COVID19 Drugs Repository: Data Collection Workflow

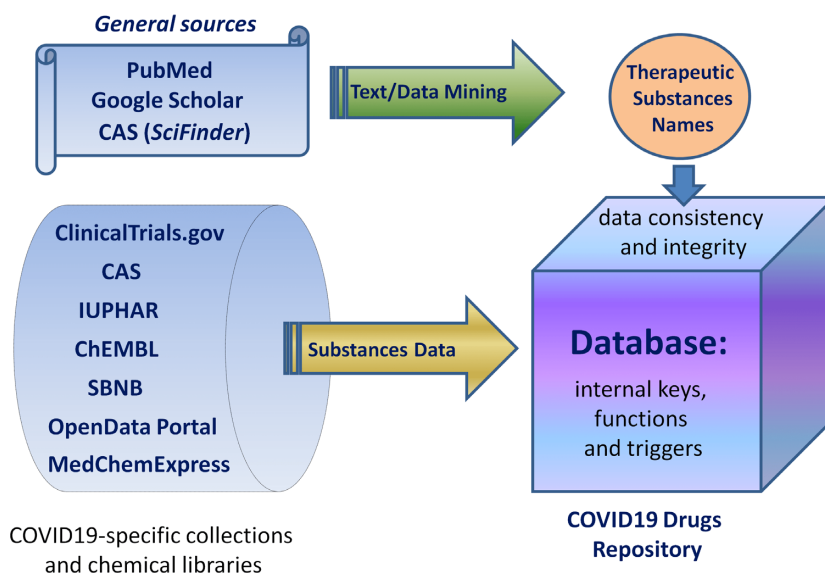


Figure 2. The data collection workflow implemented in the COVID-19 Drugs Repository.

### The database search query syntax

The database search is case-insensitive. Simple queries can include either the full or partial drug name ('Drug search' box). Advanced queries can be constructed by combining identifiers of the various databases (Table 1) and/or concepts (e.g. compound class, viral target, etc.). The 'compound class', for example, includes the following terms: «Antibody | Metabolite | Natural product | Inorganic | Peptide | Synthetic organic», while the 'viral target' category comprises the «acronym of viral name»: «BCV | BtCoV-(HKU3 | HKU5 | SHC014) | HCoV-(229E | NL63 | OC43) | (MERS|SARS)-CoV(-2)? | CCoV | FIPV | PEDV | SARS | TGEV | IBV | MHV». A query combination can be built using delimiters « and », « plus », « + », «&[]», «[]:[]» (the «[]» indicates that the 'space' delimiter may be used in the query string), for example:

*Query:* Oseltamivir; Ritonavir; DB0008934; PMID32145363; CID:37542; 118390-30-0

The search results page shows all relevant instances associated with a query drug. The HTML page is generated with hyperlinks to all databases (Table 1) associated with the drug of interest. Alternatively, the information can also be accessed by selecting a drug name from the menu in the selection field (see the Database 'Help' page: <http://covid19.md.biu.ac.il/>). The «Treatment Options» section enables access to a detailed description of the query drug.

### COVID19 Drug Repository web interface

On the menu bar, there are seven buttons (BROWSE, COLLECTIONS, DOWNLOADS, NEWS, HELP, ABOUT US and CONTACT on the right and the 'Cancer Genomics & BioComputing of complex Diseases lab' logo on the left) that users can visit with a single click. Clicking the BROWSE button returns the user to the homepage (Fig-

ure 3), which displays introductory and technical information about the Repository. Clicking the 'VIEW ALL DRUGS' button or on the 'COLLECTIONS' submenu items (e.g. 'Anti-viral drugs') brings the user to the drugs-listing webpage (Figure 3), where one can see all query drugs presented there. By entering the first letters of a query drug name in the 'Drug search' field (Figure 3), users can select their drug of interest from the dropdown menu (Figure 3).

### Features and functionality

The Repository is a COVID19-targeted collection (short-list) of ~460 items representing 184 approved drugs, 384 investigated therapeutic agents and 76 drug-like synthetic or natural chemical substances. The main focus of the repository is cross-referencing to PubMed articles linking these drugs with multiple research sources, mapping associations between the drugs and COVID19-related concepts, text/data-mining, clustering, and visualization. Furthermore, the Biopython collection of modules (44), in particular, the Entrez Bio.Entrez module (44), was implemented into our database framework so as to enable fast data retrieval via efficient command-line interactions with all NCBI resources and databases (45), sub-divided into six categories: Literature, Genes, Proteins, Genomes, Genetics, and Chemicals. A toolset of Perl and Python scripts had been created for specific tasks, such as (a) automatic generation of links between the Repository and other sources, (b) collection of data/references and creation of work tables and (c) mapping associations/concepts, with visualization options being realized via external tools.

### COVID19 drugs: search strategy

Recent information on approved drugs and therapeutic combinations thereof considered useful for the treatment



**Table 1.** Databases linked with the COVID-19 Drugs Repository

| Database Identifier  | Database/resource   | Links   |
|--|---|---|
| ACC_id (*)<br>UNII ID  | COVID19 Drug Repository<br>Unique Ingredient Identifier   | <a href="http://covid19.md.biu.ac.il/">http://covid19.md.biu.ac.il/</a><br><a href="https://www.fda.gov/industry/fda-resources-data-standards/fdas-global-substance-registration-system">https://www.fda.gov/industry/fda-resources-data-standards/fdas-global-substance-registration-system</a>  |
| CAS ID (**)<br>PubChem CID<br>PubChem SID<br>PubMed ID<br>DrugBank ID<br>IUPHAR ID<br>Clinical Trials ID | Chemical Abstracts Service<br>PubChem Compound<br>PubChem Substance<br>PubMed<br>DrugBank<br>IUPHAR/BPS<br>ClinicalTrials.gov | <a href="https://www.cas.org/">https://www.cas.org/</a><br><a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a><br><a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a><br><a href="https://pubmed.ncbi.nlm.nih.gov/">https://pubmed.ncbi.nlm.nih.gov/</a><br><a href="https://www.drugbank.ca/">https://www.drugbank.ca/</a><br><a href="https://iuphar.org/">https://iuphar.org/</a><br><a href="https://clinicaltrials.gov/">https://clinicaltrials.gov/</a> |

of SARS-CoV-2 infection has been reported at ClinicalTrials.gov and PubMed. Experimental therapeutic agents, drug-like synthetic or natural chemical substances have also been studied in the context of COVID-19, as reported in the PubMed database. To extend the coverage of our database, we reviewed articles from PubMed to extract COVID-19-related concepts and keywords. Specifically, using ‘drug repositioning’, ‘drug repurposing’, ‘COVID19’ and ‘coronavirus infections/drug therapy’ as query keywords, we reached PubMed publications with titles and/or abstracts containing combinations of these keywords. Next, we created another pattern of queries to search target information at PubMed, and using Google and Google Scholar. The pattern was expressed as a pseudo-Regular Expression (46), where text in uppercase letters denotes variable names:

‘((DRUG\_NAME)|(DRUG\_ALIAS)) ((sars)|(mers)|(corona covid)) (patients)?’

Examples of the multi-step search queries are as follows:  
‘camostat corona covid’ → [output\_1: information/references] → [therapeutic agents]

‘camostat sars’ → [output\_2: information/references] → [therapeutic agents]

‘camostat sars patients’ → [output\_3: information/references] → [therapeutic agents]

‘ONO-3403 sars patients’ → [output\_4: information/references] → [therapeutic agents]

Using this search strategy, we found ~100 substances with activities associated with COVID-19. The queries and the patterns used, and the information obtained daily are the main sources for Repository updates.

All ‘active’ chemical entities/substances (i.e. those with demonstrated or proposed therapeutic potential) were collected and linked via their identifiers in CAS, PubChem, IUPHAR, etc. Furthermore, we adopted the web-based text clustering engine Carrot<sup>2</sup> (47) for visualization of pair-wise ‘drug-COVID-19 concept’ associations found in PubMed abstracts for each pair.

In this version of our database, COVID19 drugs were mapped to a dictionary of 21 terms related to concepts of ‘COVID-19’, e.g. ‘viral infections’, ‘respiratory diseases’, ‘inflammatory cell’, ‘coronavirus pneumonia’, etc. (Supplemental Table S1). These terms are the most frequent words/combinations clustered around the central words such as ‘virus’, ‘infection’, ‘inflammation’, ‘pneumonia’, ‘lungs’. To create the concepts’ dictionary, a variety of clusters were generated by experimentation with different hierarchical clustering algorithms applied to the collection of PubMed titles/abstracts. Links to all PubMed

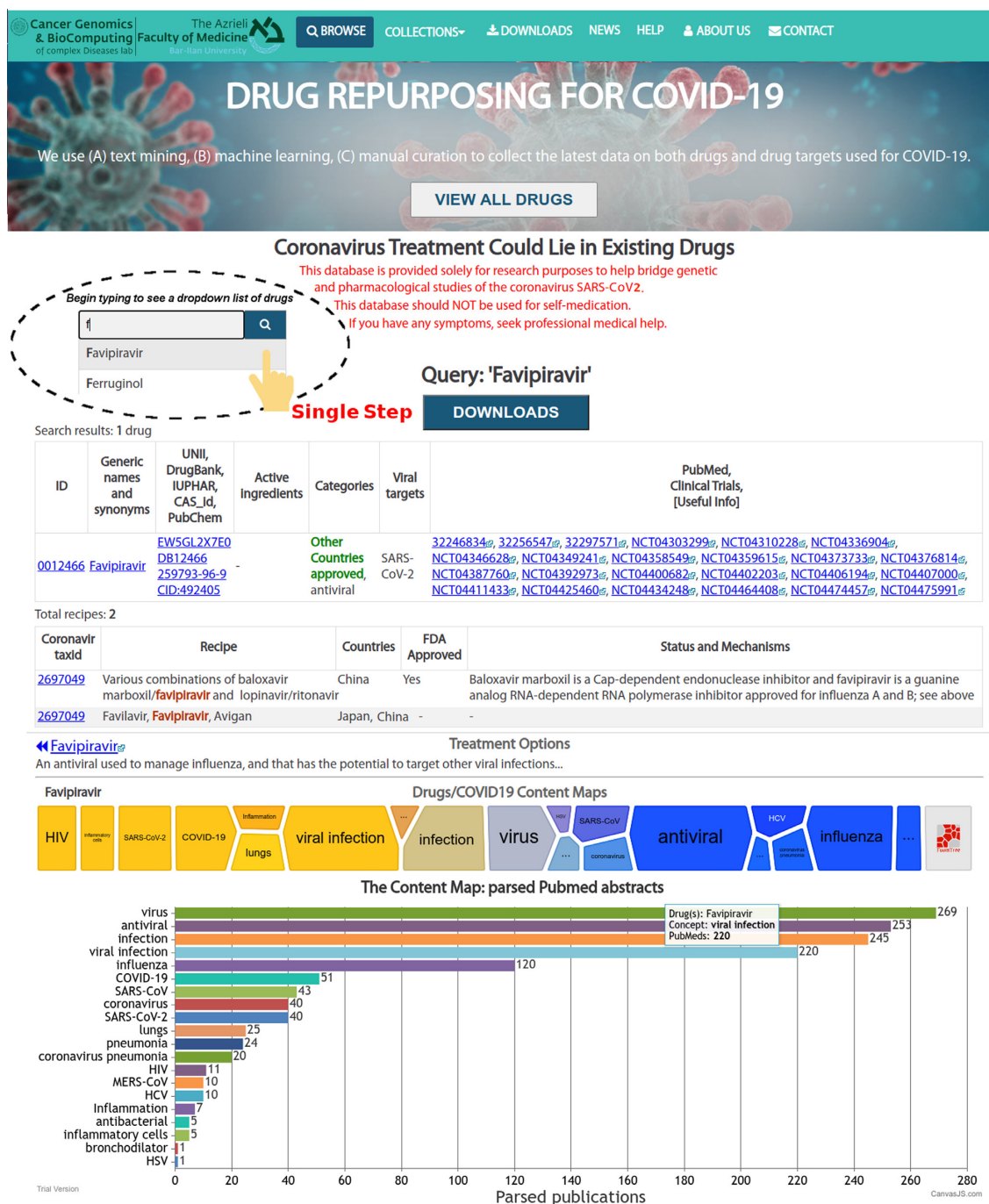
abstracts associated with these ‘drug-concept’ pairs were generated and enumerated using Python scripts. PubMed search queries were created according to PubMed query syntax (48) and MeSH terms (49). The automatically generated tables (available in the ‘DOWNLOADS’ section) list the number of retrieved PubMed publications corresponding to each ‘drug-COVID-19 concept’ pair. These numbers are hyperlinked with the corresponding PubMed publications. Links to references and the Carrot<sup>2</sup> text clustering and visualization tool can be updated on a regular basis. Such updates are necessary as the web and PubMed database are constantly expanding, with new references and sites appearing daily. All desired data can be downloaded from the COVID19 Drug Repository website (link) as Excel tables containing the list of keywords (i.e., the ‘dictionary’) used for text-mining and mapping. In subsequent versions of the database, users will be able to modify the list or introduce additional concepts. With this simple mapping tool, one can discover and visualize new concepts and associations that would not otherwise be found.

### Drug repurposing sources

Currently, there are 384 mapped drug names mentioned in 960 COVID-19 clinical studies (data retrieved on August 15, 2020), with at least 1 drug intervention (Table 2). None of these drugs are novel. Rather, they exemplify a ‘drug repurposing/repositioning’ approach (26,34,50–52). Recently, numerous COVID19-specific web pages and chemical libraries have been created by different research organizations (CAS, IUPHAR (53,54), ChEMBL, Open-Data Portal (<https://opendata.ncats.nih.gov/covid19/index.html>), etc.) and companies (MedChem Express), and used for the high throughput screening against SARS-CoV-2 infection (55–63). All these molecular libraries and collections (Figure 2, Table 2) are being used in our data collection process (Figure 2), and listed in the Repository web page (‘Useful Links’).

### Drug-gene interaction networks

The Biopython/Entrez-based Python command-line script (as discussed in the Features and functionality section) was created to access the NCBI Gene database (45), and to automatically retrieve human or microbial (and in particular, viral) genes associated with a given list of drugs or chemical substances. The output (Figure 4) provides a list of genes with a short description of the biological role associated with each gene product in the output list.



**Figure 3.** Web interface of the COVID19 Drug Repository. Selection of the drug of interest (e.g. *Favipiravir*) from the dropdown menu/list.

Those genes associated with a set of drugs can be analysed, clustered, or served as input for building 'drug-gene' networks and then visualized using external programs. As a working example, protein-protein association networks were built for output gene sets using the STRINGv11 database (64). Moreover, the application programming interface (API) implemented in the STRINGv11 database enables efficient interaction of external databases with the STRING visualization and analysis tools (64). For example, visualization of the set of genes associated with the

PDE5A inhibitor sildenafil, a vasodilator agent, revealed other interesting targets (Figure 4), such as the enzyme PDE6G. Both enzymes are active in the lungs (65,66). Further network analysis of available data showed that PDE5A/PDE6G inhibition by sildenafil in lung blood vessels can trigger different anti-inflammatory pathways.

To build a network (Figure 5), we extracted additional information from the literature and external databases. In the PDE5A and PDE6G protein expression summaries obtained from the Human Protein Atlas (67), the PGE6G

**Table 2.** COVID19-specific collections and chemical libraries

| Resource   | Dataset (title, catalogue, or category)        | Drugs, active agents                            | Links   |
|--|--|---|---|
| ClinicalTrials.gov   | COVID-19 studies by mapped drug intervention   | 384   | <a href="https://clinicaltrials.gov/ct2/covid_view/drugs">https://clinicaltrials.gov/ct2/covid_view/drugs</a>   |
| CAS  | COVID-19 Antiviral Candidate Compounds Dataset | ~50,000   | <a href="https://www.cas.org/covid-19-antiviral-compounds-dataset">https://www.cas.org/covid-19-antiviral-compounds-dataset</a>   |
| IUPHAR   | Ligands relevant to SARS-CoV-2(COVID-19)       | 64 items  | <a href="https://www.guidetopharmacology.org/GRAC/CoronavirusForward">https://www.guidetopharmacology.org/GRAC/CoronavirusForward</a>                                   |
| ChEMBL   | ChEMBL27 SARS CoV-2 Release, 8 datasets        | 133 selected (IC50/EC50 better than 10 $\mu$ M) | <a href="http://chembl.blogspot.com/2020/05/chembl27-sars-cov-2-release.html">http://chembl.blogspot.com/2020/05/chembl27-sars-cov-2-release.html</a>                   |
| Structural Bioinformatics and Network Biology Group (SBNB) OpenData Portal | Drugs from sthe COVID19 literature             | 307   | <a href="https://sbnb.irbbarcelona.org/covid19/">https://sbnb.irbbarcelona.org/covid19/</a>   |
| MedChemExpress   | NCATS Anti-infectives Collection               | 740   | <a href="https://opendata.ncats.nih.gov/covid19/index.html">https://opendata.ncats.nih.gov/covid19/index.html</a>   |
| MedChemExpress   | SARS-CoV List of Drugs                         | 76  | <a href="https://www.medchemexpress.com/Targets/SARS-CoV.html">https://www.medchemexpress.com/Targets/SARS-CoV.html</a>   |
| MedChemExpress   | Anti-Virus Compound Library                    | 512   | <a href="https://www.medchemexpress.com/screening/Anti-virus_Compound_Library.html">https://www.medchemexpress.com/screening/Anti-virus_Compound_Library.html</a>       |
| MedChemExpress   | Anti-COVID-19 Compound Library                 | 1552  | <a href="https://www.medchemexpress.com/screening/anti-covid-19-compound-library.html">https://www.medchemexpress.com/screening/anti-covid-19-compound-library.html</a> |

```

C:\Users\user>python 1234_Gene_i-Link2NumPubMed.py
"sildenafil" AND "Homo sapiens"[porgn] AND alive[prop]
  0          1          2
0  4846    NOS3          nitric oxide synthase 3
1  5290    PIK3CA  phosphatidylinositol-4,5-bisphosphate 3-kinase...
2  1588    CYP19A1  cytochrome P450 family 19 subfamily A member 1
3  2784    GNB3          G protein subunit beta 3
4   948    CD36          CD36 molecule
5 10599    SLC01B1  solute carrier organic anion transporter famil...
6   8654    PDE5A          phosphodiesterase 5A
7 28234    SLC01B3  solute carrier organic anion transporter famil...
8   5148    PDE6G          phosphodiesterase 6G
C:\Users\user>

```

**Figure 4.** Example output of the *DruGeNetwork* python script: list of genes (column '1') with a short description/annotation of the underlying biological mechanism associated with each gene (column '1'). The output is generated for the query keywords sent to the NCBI Gene resources: "sildenafil" AND "Homo sapiens"[porgn] AND alive[prop].

gene is categorized as 'Group enriched' in natural killer (NK) cells, according to consensus transcriptomics data. NK cells, acting as cytotoxic lymphocytes, are involved in innate immune system regulation, including rapid cytokine production in the presence of virus-infected cells (67). The NK-mediated antiviral immune response is associated with the NCR1 gene that encodes the natural cytotoxicity receptor 1 (68). In the next step, both the PDE6 and NCR1 genes were detected in the Chronic Obstructive Pulmonary Disease (COPD)-related Gene Set using Harmonizome on the collection of 'omics' Big Data sets (69). This gene set was deposited in the GEO Signatures of Differentially Expressed Genes for Diseases, under the name 'COPD-Chronic Obstructive Pulmonary Disease.Muscle-Striated (Skeletal)-Diaphragm (MMHCC)\_GSE47'. The data show that the expression of the PDE6G gene is significantly increased, whereas decreased expression was reported for the NCR1 gene.

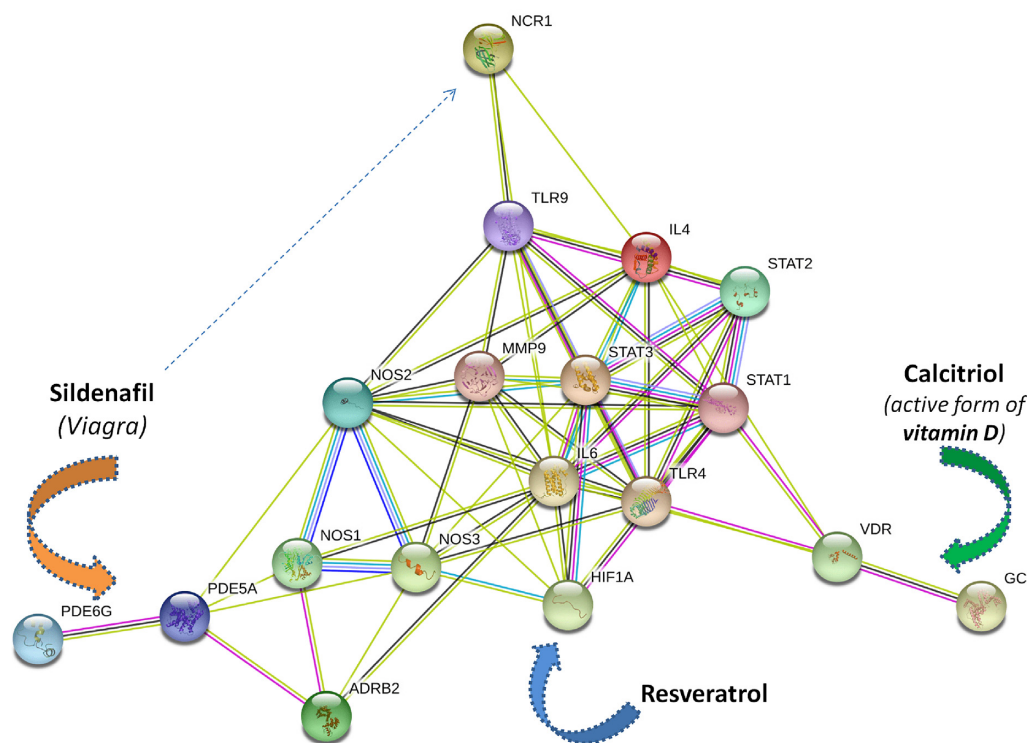
Therefore, in the context of drug repurposing strategies, it is reasonable to expect that sildenafil will be use-

ful for the treatment of COVID19 complications. Accordingly, two recent clinical studies (ClinicalTrials.gov identifiers NCT04304313 and NCT04489446) were initiated to study the efficacy and safety of sildenafil in patients with COVID-19 (NCT04304313), and to assess the role of sildenafil in improving oxygenation among hospitalised patients (NCT04489446).

### Identification of drug target genes

We extracted target gene information for each putative COVID-19 drug from the Therapeutic Target Database (70), and by text-mining of the literature at PubMed. To understand the expression profile of drug target genes identified in this manner in the COVID-19 infection, we performed transcriptome analysis of infected bronchial epithelial cells. For this, we retrieved raw RNA-sequencing data for SARS-CoV-2-infected bronchial epithelial cells from the sequence read archive (SRA) database under accession no. PRJNA615032 (71). The FASTQ files were mapped and





**Figure 5.** The Drug-Gene local Network built using the output for sildenafil (A). Further functional enrichment reveals the group of genes involved in the innate immunity and inflammatory processes. Moreover, VDR (receptors activated by calcitriol), and the HIF1A-STAT3 path suppressed by resveratrol, are also involved in the local anti-inflammatory pharmacological network, thereby suggesting the “sildenafil-resveratrol-vitamin D3” drug combination to treat the COVID-19 complications

aligned to the hg38 reference genome using STAR (72). Differentially expressed genes were identified using edgeR (73), with parameters set at 2.0-fold change and  $<0.05$   $P$ -value cut-off. We thus found target genes for 41 drugs from our database (e.g. sildenafil as discussed in previous paragraph) which are significantly differentially expressed during COVID-19 infection. These drug-gene pairs are given in the Supplemental Table S2.

#### COVID19 Drug Repository server, hardware and software requirements

The COVID19 Drug Repository server was built on an Apache web server and deployed on the RedHat Enterprise Linux (RHEL) 7.4 server of an Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz and 32GB RAM unit. The COVID19 Drug Repository has a code base and infrastructure similar to that of the ChiTaRS database (74,75). The COVID19 Drug Repository website is compatible with modern web browsers (such as Chrome, Firefox, Microsoft Edge, Opera and Safari), provided that JavaScript is enabled. We recommend using the latest release version of these web browsers for optimal rendering.

#### Downloads

The COVID19 Drug Repository not only provides extended ‘Search’ options but also offers the possibility to download all database tables and data sets in a user-friendly manner. The repository is available at: <http://covid19.md.biu.ac.il/>.

## CONCLUSIONS AND FUTURE PLANS

The COVID19 Drug Repository maps data from chemogenomics and pharmacogenomics studies and provides viral and human genomics and proteomics information on approved drugs and other therapeutics. The database enables the user to focus on different levels of complexity, starting from general information, clinical trials and formulations, and increasing the resolution to the level of molecular mechanisms of drug action. Therefore, the database can serve as a navigation and recommendation tool both for research and for healthcare purposes. Future plans include the following additions to the database: (a) continuous updating with new data on approved drugs, experimental drugs, and drug-like synthetic or natural chemical substances; (b) automatic machine learning and text-mining-based annotation and visualization of ‘Mode of Action’ (MoA) data, as well as ‘Drug-Gene’, and ‘Drug-Symptom’ networks; and (c) incorporation of the Drugs/NGS analysis tools (‘transcriptomics’) to accelerate the translation of knowledge for use as personalized medicine for COVID19 patients.

#### DATA AVAILABILITY

<http://covid19.md.biu.ac.il/>.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

E.L. has been working as a volunteer at the Frenkel-Morgenstern's lab in the COVID-19 related project.

## FUNDING

M.F-M is supported by the Kamin grant of Israel Innovation Authority (#66824, 2019-2021) and COVID-19 Data Science Institute (DSI) grant, Bar-Ilan University (#247017, 2020). SM is supported by the PBC fellowship for outstanding postdoctoral researchers from India by the Council of Higher Education, Israel (VaTaT #104-2019, at M. F-M lab).

*Conflict of interest statement.* Authors declare no conflict of interest.

## REFERENCES

- Chen, Q., Allot, A. and Lu, Z. (2020) Keep up with the latest coronavirus research. *Nature*, **579**, 193.
- Lu Wang, L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., Funk, K., Kinney, R., Liu, Z., Merrill, W. *et al.* (2020) COVID-19: The Covid-19 Open Research Dataset. arXiv doi: <https://arxiv.org/abs/2004.10706v2>, 10 July 2020, preprint: not peer reviewed.
- Swanson, D.R. (2008) In: Bruza, P and Weeber, M (eds). *Literature-based Discovery*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–11.
- Wei, C.H., Kao, H.Y. and Lu, Z. (2013) PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res.*, **41**, W518–W522.
- Roberts, K., Alam, T., Bedrick, S., Demner-Fushman, D., Lo, K., Soboroff, I., Voorhees, E., Wang, L.L. and Hersh, W.R. (2020) TREC-COVID: rationale and Structure of an Information Retrieval Shared Task for COVID-19. *J. Am. Med. Inform. Assoc.*, **27**, 1431–1436.
- Crichton, G., Baker, S., Guo, Y. and Korhonen, A. (2020) Neural networks for open and closed Literature-based Discovery. *PLoS One*, **15**, e0232891.
- Zhang, E., Gupta, N., Nogueira, R., Cho, K. and Lin, J. (2020) Rapidly deploying a neural search engine for the covid-19 open research dataset: Preliminary thoughts and lessons learned. arXiv doi: <https://arxiv.org/abs/2004.05125>, 10 April 2020, preprint: not peer reviewed.
- Tarasova, O., Ivanov, S., Filimonov, D.A. and Poroikov, V. (2020) Data and text mining help identify key proteins involved in the molecular mechanisms shared by SARS-CoV-2 and HIV-1. *Molecules*, **25**, 2944.
- Wei, C.H., Harris, B.R., Li, D., Berardini, T.Z., Huala, E., Kao, H.Y. and Lu, Z. (2012) Accelerating literature curation with text-mining tools: a case study of using PubTator to curate genes in PubMed abstracts. *Database (Oxford)*, **2012**, bas041.
- Wei, C.H., Allot, A., Leaman, R. and Lu, Z. (2019) PubTator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Res.*, **47**, W587–W593.
- Osiński, S. and Weiss, D. (2004) Conceptual Clustering Using Lingo Algorithm: Evaluation on Open Directory Project Data. In: Kłopotek, M.A., Wierzchoń, S.T. and Trojanowski, K. (eds). *Intelligent Information Processing and Web Mining. Advances in Soft Computing*. Springer, Berlin, Heidelberg. Vol: **25**, pp. 369–377.
- Tagore, S., Gorohovski, A., Jensen, L.J. and Frenkel-Morgenstern, M. (2019) ProtFus: a comprehensive method characterizing protein-protein interactions of fusion proteins. *PLoS Comput. Biol.*, **15**, e1007239.
- Meyer, U.A., Zanger, U.M. and Schwab, M. (2013) Omics and drug response. *Annu. Rev. Pharmacol. Toxicol.*, **53**, 475–502.
- Zheng, X.F. and Chan, T.F. (2002) Chemical genomics: a systematic approach in biological research and drug discovery. *Curr. Issues Mol. Biol.*, **4**, 33–43.
- Engelberg, A. (2004) Iconix Pharmaceuticals, Inc.—removing barriers to efficient drug discovery through chemogenomics. *Pharmacogenomics*, **5**, 741–744.
- Magariños, M.P., Carmona, S.J., Crowther, G.J., Ralph, S.A., Roos, D.S., Shanmugam, D., Van Voorhis, W.C. and Agüero, F. (2012) TDR targets: a chemogenomics resource for neglected diseases. *Nucleic Acids Res.*, **40**, D1118–D1127.
- Wang, L., Ma, C., Wipf, P., Liu, H., Su, W. and Xie, X.Q. (2013) TargetHunter: an in silico target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *AAPS J.*, **15**, 395–406.
- Tribut, O., Lessard, Y., Reymann, J.M., Allain, H. and Bentué-Ferrer, D. (2002) Pharmacogenomics. *Med. Sci. Monit.*, **8**, RA152–RA163.
- Hood, E. (2003) Pharmacogenomics: the promise of personalized medicine. *Environ. Health Perspect.*, **111**, A581–A589.
- Weinshilboum, R. and Wang, L. (2004) Pharmacogenomics: bench to bedside. *Nat. Rev. Drug Discov.*, **3**, 739–748.
- Service, R.F. (2005) Pharmacogenomics. Going from genome to pill. *Science*, **308**, 1858–1860.
- Blake, C.A. and Sobel, B.E. (2008) Progress in pharmacogenomics and its promise for medicine. *Exp. Biol. Med. (Maywood)*, **233**, 1482–1483.
- Wang, L. (2010) Pharmacogenomics: a systems approach. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **2**, 3–22.
- Burt, T. and Dhillon, S. (2013) Pharmacogenomics in early-phase clinical development. *Pharmacogenomics*, **14**, 1085–1097.
- Musa, A., Tripathi, S., Dehmer, M., Yli-Harja, O., Kauffman, S.A. and Emmert-Streib, F. (2019) Systems pharmacogenomic landscape of drug similarities from LINCS data: Drug Association Networks. *Sci. Rep.*, **9**, 7849.
- Kalamara, A., Tobalina, L. and Saez-Rodriguez, J. (2018) How to find the right drug for each patient? Advances and challenges in pharmacogenomics. *Curr. Opin. Syst. Biol.*, **10**, 53–62.
- Kirk, R.J., Hung, J.L., Horner, S.R. and Perez, J.T. (2008) Implications of pharmacogenomics for drug development. *Exp. Biol. Med. (Maywood)*, **233**, 1484–1497.
- Sharifi-Noghabi, H., Zolotareva, O., Collins, C.C. and Ester, M. (2019) MOLI: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics*, **35**, i501–i509.
- Pulley, J.M., Rhoads, J.P., Jerome, R.N., Challa, A.P., Erreger, K.B., Joly, M.M., Lavieri, R.R., Perry, K.E., Zaleski, N.M., Shirey-Rice, J.K. *et al.* (2020) Using what we already have: uncovering new drug repurposing strategies in existing omics data. *Annu. Rev. Pharmacol. Toxicol.*, **60**, 333–352.
- Sindelar, R.D. (2013) Genomics, other “Omic” technologies, personalized medicine, and additional biotechnology-related techniques. In: Crommelin, D.J.A., Sindelar, R.D. and Meibohm, B. (eds). *Pharmaceutical Biotechnology: Fundamentals and Applications*. Springer International Publishing, NY, pp. 179–221.
- Urán Landaburu, L., Berenstein, A.J., Videla, S., Maru, P., Shanmugam, D., Chernomoretz, A. and Agüero, F. (2020) TDR Targets 6: driving drug discovery for human pathogens through intensive chemogenomic data integration. *Nucleic Acids Res.*, **48**, D992–D1005.
- Feng, Z., Chen, M., Liang, T., Shen, M., Chen, H. and Xie, X.Q. (2020) Virus-CKB: an integrated bioinformatics platform and analysis resource for COVID-19 research. *Brief. Bioinform.*, **bbaa155**.
- Duran-Frigola, M., Pauls, E., Guitart-Pla, O., Bertoni, M., Alcalde, V., Amat, D., Juan-Blanco, T. and Aloy, P. (2020) Extending the small-molecule similarity principle to all levels of biology with the chemical checker. *Nature Biotechnology*, **38**, 1087–1096.
- Singh, N., Decroly, E., Khatib, A.M. and Villoutreix, B.O. (2020) Structure-based drug repositioning over the human TMPRSS2 protease domain: search for chemical probes able to repress SARS-CoV-2 Spike protein cleavages. *Eur. J. Pharm. Sci.*, **153**, 105495.
- Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O’Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L. *et al.* (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, **583**, 459–468.
- Takahashi, T., Luzum, J.A., Nicol, M.R. and Jacobson, P.A. (2020) Pharmacogenomics of COVID-19 therapies. *NPJ Genom. Med.*, **5**, 35.
- Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
- Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z. *et al.* (2018)



- DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
39. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B. *et al.* (2019) PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.*, **47**, D1102–D1109.
  40. Armstrong, J.F., Faccenda, E., Harding, S.D., Pawson, A.J., Southan, C., Sharman, J.L., Campo, B., Cavanagh, D.R., Alexander, S.P.H., Davenport, A.P. *et al.* (2020) The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.*, **48**, D1006–D1021.
  41. Huffmanberger, M.A. and Wington, R.L. (1975) Chemical Abstracts Service approach to management of large data bases. *J. Chem. Inf. Comput. Sci.*, **15**, 43–47.
  42. Weisgerber, D.W. (1997) Chemical Abstracts Service Chemical Registry System: history, scope, and impacts. *J. Am. Soc. Inf. Sci.*, **48**, 349–360.
  43. Widenius, M., Axmark, D. and Arno, K. (2002) In: *MySQL Reference Manual: Documentation From the Source*. O'Reilly Media, Inc.
  44. Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
  45. Coordinators, N.R. (2016) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **44**, D7–D19.
  46. Thompson, K. (1968) Programming techniques: regular expression search algorithm. *Commun. ACM*, **11**, 419–422.
  47. Stefanowski, J. and Weiss, D. (2003) Carrot2 and Language Properties in Web Search Results Clustering. In: Menasalvas, E., Segovia, J. and Szczepaniak, P.S. (eds) *Advances in Web Intelligence*. Springer, Berlin, Heidelberg, pp. 240–249.
  48. Sayers, E. (2010) E-utilities Quick Start. *Entrez Programming Utilities Help [Internet]*. Bethesda, MD: National Center for Biotechnology Information (US). <http://www.ncbi.nlm.nih.gov/books/NBK25500>.
  49. Lipscomb, C.E. (2000) Medical subject headings (MeSH). *Bull. Med. Libr. Assoc.*, **88**, 265.
  50. Sahu, N.U. and Kharkar, P.S. (2016) Computational Drug Repositioning: a lateral approach to traditional drug discovery? *Curr. Top. Med. Chem.*, **16**, 2069–2077.
  51. Sam, E. and Athri, P. (2019) Web-based drug repurposing tools: a survey. *Brief. Bioinform.*, **20**, 299–316.
  52. Donner, Y., Kazmierczak, S. and Fortney, K. (2018) Drug repurposing using deep embeddings of gene expression profiles. *Mol. Pharm.*, **15**, 4314–4325.
  53. Alexander, S.P.H., Armstrong, J.F., Davenport, A.P., Davies, J.A., Faccenda, E., Harding, S.D., Levi-Schaffer, F., Maguire, J.J., Pawson, A.J., Southan, C. *et al.* (2020) A rational roadmap for SARS-CoV-2/COVID-19 pharmacotherapeutic research and development: IUPHAR Review 29. *Br. J. Pharmacol.*, **177**, 4942–4966.
  54. Faccenda, E., Armstrong, J., Davenport, A., Harding, S., Pawson, A., Southan, C. and Davies, J. (2020) Coronavirus Information. IUPHAR/BPS Guide to Pharmacology, <https://www.guidetopharmacology.org/coronavirus.jsp>.
  55. Jeon, S., Ko, M., Lee, J., Choi, I., Byun, S.Y., Park, S., Shum, D. and Kim, S. (2020) Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs. *Antimicrob. Agents Chemother.*, **64**, e00819-20.
  56. Riva, L., Yuan, S., Yin, X., Martin-Sancho, L., Matsunaga, N., Burgstaller-Muehlbacher, S., Pache, L., De Jesus, P.P., Hull, M.V., Chang, M. *et al.* (2020) A large-scale drug repositioning survey for SARS-CoV-2 antivirals. bioRxiv doi: <https://doi.org/10.1101/2020.04.16.044016>, 17 April 2020, preprint: not peer reviewed.
  57. Brimacombe, K.R., Zhao, T., Eastman, R.T., Hu, X., Wang, K., Backus, M., Baljinyam, B., Chen, C.Z., Chen, L., Eicher, T. *et al.* (2020) An OpenData portal to share COVID-19 drug repurposing data in real time. bioRxiv doi: <https://doi.org/10.1101/2020.06.04.135046>, 05 June 2020, preprint: not peer reviewed.
  58. Mirabelli, C., Wotring, J.W., Zhang, C.J., McCarty, S.M., Fursmidt, R., Frum, T., Kadambi, N.S., Amin, A.T., O'Meara, T.R., Pretto, C.D. *et al.* (2020) Morphological cell profiling of SARS-CoV-2 infection identifies drug repurposing candidates for COVID-19. bioRxiv doi: <https://doi.org/10.1101/2020.05.27.117184>, 27 May 2020, preprint: not peer reviewed.
  59. Tang, H., Abouleila, Y., Si, L., Ortega-Prieto, A.M., Mummery, C.L., Ingber, D.E. and Mashaghi, A. (2020) Human organs-on-chips for virology. *Trends Microbiol.*, **28**, 934–946.
  60. Weston, S., Coleman, C.M., Haupt, R., Logue, J., Matthews, K., Li, Y., Reyes, H.M., Weiss, S.R. and Frieman, M.B. (2020) Broad anti-coronaviral activity of FDA approved drugs against SARS-CoV-2. *J. Virol.*, **94**, e01218-20.
  61. Touret, F., Gilles, M., Barral, K., Nougairède, A., van Helden, J., Decroly, E., de Lamballerie, X. and Coutard, B. (2020) In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. *Sci. Rep.*, **10**, 13093.
  62. Ellinger, B., Bojkova, D., Zaliani, A., Cinatl, J., Clausen, C., Westhaus, S., Reinshagen, J., Kuzikov, M., Wolf, M. and Geisslinger, G. (2020) Identification of inhibitors of SARS-CoV-2 in-vitro cellular toxicity in human (Caco-2) cells using a large scale drug repurposing collection. *ResearchSquare*, doi:10.21203/rs.3.rs-23951/v1.
  63. Heiser, K., McLean, P.F., Davis, C.T., Fogelson, B., Gordon, H.B., Jacobson, P., Hurst, B., Miller, B., Alfa, R.W., Earnshaw, B.A. *et al.* (2020) Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of human cells infected with SARS-CoV-2. bioRxiv doi: <https://doi.org/10.1101/2020.04.21.054387>, 23 April 2020, preprint: not peer reviewed.
  64. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P. *et al.* (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, D607–D613.
  65. Nikolova, S., Guenther, A., Savai, R., Weissmann, N., Ghofrani, H.A., Konigshoff, M., Eickelberg, O., Klepetko, W., Voswinckel, R., Seeger, W. *et al.* (2010) Phosphodiesterase 6 subunits are expressed and altered in idiopathic pulmonary fibrosis. *Respir. Res.*, **11**, 146.
  66. Hemnes, A.R., Zaiman, A. and Champion, H.C. (2008) PDE5A inhibition attenuates bleomycin-induced pulmonary fibrosis and pulmonary hypertension through inhibition of ROS generation and RhoA/Rho kinase activation. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, **294**, L24–L33.
  67. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A. *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
  68. Pessino, A., Sivori, S., Bottino, C., Malaspina, A., Morelli, L., Moretta, L., Biassoni, R. and Moretta, A. (1998) Molecular cloning of NKp46: a novel member of the immunoglobulin superfamily involved in triggering of natural cytotoxicity. *J. Exp. Med.*, **188**, 953–960.
  69. Rouillard, A.D., Gunderen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G. and Ma'ayan, A. (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)*, **2016**, baw100.
  70. Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., Zhang, R., Zhu, J., Ren, Y., Tan, Y. *et al.* (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.*, **48**, D1031–D1041.
  71. Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.C., Uhl, S., Hoagland, D., Möller, R., Jordan, T.X., Oishi, K., Panis, M., Sachs, D. *et al.* (2020) Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*, **181**, 1036–1045.
  72. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
  73. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
  74. Balamurali, D., Gorohovski, A., Detroja, R., Palande, V., Raviv-Shay, D. and Frenkel-Morgenstern, M. (2019) ChiTaRS 5.0: the comprehensive database of chimeric transcripts matched with druggable fusions and 3D chromatin maps. *Nucleic Acids Res.*, **48**, D825–D834.
  75. Frenkel-Morgenstern, M., Gorohovski, A., Lacroix, V., Rogers, M., Ibanez, K., Boulosa, C., Andres Leon, E., Ben-Hur, A. and Valencia, A. (2013) ChiTaRS: a database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data. *Nucleic Acids Res.*, **41**, D142.