COV-Drug Target interaction server for Covid-19 Drug Repurposing

Kamal Rawal[#1], Prashant Singh[1], Robin Sinha[1], Priya Kumari[1], Swarsat Kaushik Nath[1], Ridhima[1], Sukriti Sahai[1], Sweety Dattatraya Shinde[1], Nikita Garg[1], Preeti P.[1], Trapti Sharma[1]

1. Amity Institute of Biotechnology, Amity University, Uttar Pradesh, India

#Corresponding Author
Email ID: kamal.rawal@gmail.com
Centre for Computational Biology and Bioinformatics, AIB
Amity University, Noida.

Keywords: Bioinformatics, drug repurposing, artificial intelligence, COVID-19, molecular targets

Supplementary Data Website: https://tinyurl.com/drugX-Drug-target

COV-DRUGX Software Pipeline : http://drugx.kamalrawal.in/cov_target/

Abstract

The outbreak of the novel coronavirus disease COVID-19, caused by the SARS-CoV-2 virus has killed over 5 million people to date. So, there is an urgent requirement for new and effective medications that can treat the disease caused by SARS-CoV-2. To find new drugs, identification of drug targets is necessary (Chen et al., 2016). Number of research studies have identified therapeutic targets such as helicases, transmembrane serine protease 2, cathepsin L, cyclin G-associated kinase, adaptor-associated kinase 1, two-pore channel, viral virulence factors, 3-chymotrypsin-like protease, suppression of excessive inflammatory response, inhibition of viral membrane, nucleocapsid, envelope, and accessory proteins, and inhibition of endocytosis. Here we present a web enabled tool which helps in ranking the COVID-19 drugs based upon underlying molecular targets. The users are allowed to give drugs in SMILE format and the tools will provide the list of relevant targets related to COVID-19.

1. Introduction

A functionally altered protein target plays a key role in the disease and a drug treats the disease by inhibiting or activating the target. Thus, drug repositioning can act on each of these three levels: disease, target, or drug [**Parisi et al., 2020**]. Many drugs are being developed and clinically tested for treating COVID-19. The knowledge of the targets and the properties of the drugs are highly useful for facilitating drug repurposing, clinical evaluation, and drug and target discovery efforts. A variety of databases have provided dedicated information sources and access facilities to support these efforts [**Wang et al., 2020**]. In drug-repurposing, the knowledge of the drug targets relevant to COVID-19 therapeutics is key for clinical/biologic evaluations of drug efficacies, investigations of therapeutic mechanisms, and searches of drug-repurposing opportunities [**Guy et al., 2020; Zhang et al., 2020**]. Many drugs have emerged from the drug-repurposing efforts for suppressing the post-infection disease progression and life-threatening symptoms, and new targets of high drug-repurposing potential are discovered by investigating virus-host interaction and infection-induced host proteomics change [**Guy et al., 2020**].

Despite some successes, drug repurposing remains a challenge for two main reasons: (1) validating druggable therapeutic target(s) associated with the disease, and (2) confidently establishing the repertoire of protein target interactions for the FDA approved drug set [**T Issa et al., 2015**]. A variety of methods for establishing drug-target interactions are employed in both academia and industry. High-throughput screening (HTS) strategies are used for establishing interactions for large drug libraries against protein targets of interest [**Fox et al., 1999**]. The amount of potential druggable disease-related targets is also exponentially increasing [**Griffith et al., 2013**] along with the number of synthesizable drugs [**Irwin et al., 2012**]. Creating the vast possible drug-target space of true interactions and further narrowing it to that of physiologic- and disease-relevance remains a great challenge. Many efforts for computationally predicting drug-target interactions exist, spanning both chemo-centric [**Keiser et al., 2007; Warner et al., 2012**] and target-based methodologies [**Bolton et al., 2011; Chen et al., 2001**]. The target-based strategy is based on docking whereas chemo-centric uses chemical and physical information of the drug.

Disease-centric repositioning, as we define it, consists of the re-profiling of drugs among different types of a disease, such as two types of cancer. The underlying assumption for disease-centric repositioning is that different types of diseases share similar guiding principles. For example, the tyrosine-protein kinase ABL has recently been suggested as a novel target in Parkinson's disease [**Kellenberger et al., 2008**]. Hence, its inhibitors, such as nilotinib, might be effective against this syndrome [**Kellenberger et al., 2008**]. This indication shift from cancer to

neurodegeneration is driven by the target tyrosine-protein kinase ABL and represents a case of target-centric repositioning.

Drug-centric repositioning occurs when a novel target connected to a certain indication is predicted for a given drug. For example, valproic acid is for bipolar disorder and seizures because of its ability to bind to the mitochondrial enzymes succinate-semialdehyde dehydrogenase (ALDH5A1) and 4-aminobutyrate aminotransferase (ABAT).

Previously, we have developed several machine learning and bioinformatics platforms. These include, text mining and network biology based systems [**Jagannadham et al., 2016**], vaccine discovery system [**Rawal et al. 2021**], next generation sequencing analysis system for cancer genomes [**Preeti et al., 2021, Rawal et al 2011**]. Here we present a web enabled system for scoring the drug targets for COVID-19 related drugs. The module fetches the drug target from the experimental drug-protein target interactions obtained from the Therapeutic Target Database (TTD) and presents useful data for drug repurposing activities.

2. Implementation

The COV-Drug-Target is a tool written in python to screen the associated targets of a particular drug. We have included 376 drug targets from humans which have been reported to play a role in the pathophysiology of the COVID (**Supplementary Table 1**). The targets were collected using the drugs displaying some functional role in the treatment of COVID-19 (positive drug dataset). The positive drug dataset was collected using a hybrid approach consisting of text mining and deep curation approach utilizing the keyword "**drug repurposing and covid 19**". A total of 24,800 abstracts were collected till the date (17.05.2021). Out of these abstracts, we collected 393 drugs (**Supplementary Table 2**). Each drug was curated with literature evidence (**Supplementary Table 3**). Out of 393, we extracted 261 unique drugs with their SMILE notations (**Supplementary Table 4**). Further, we screened Therapeutic Target Database (TTD) [**Zhou et al., 2021**] and extracted drug targets related to those members of the positive datasets.

Next, we have extracted 32,356 drug-targets information from the TTD (http://db.idrblab.net/ttd/). After the removal of duplicates, 31,356 unique drug-targets were collected (**Supplementary Table 5**). The schematic workflow of the drug target module is shown in **Supplementary Figure 1.**

3. Usage

The module provides the target for a given drug name that is associated with COVID-19. Users are allowed to provide either drug names (separated by pipe in a text file) or SMILE notation of the drugs (separated by a newline character in a text file). The module accepts drug names as a query and it searches for its target in the TTD datasets. The module picks the target of the provided drug and moves towards the COVID-19 target datasets. If the module got the target name in the COVID-19 dataset which was obtained from the TTD dataset then the module predicts a score of 1 otherwise it will give 0

4. Result and Discussion

As a case study, we have collected three drug datasets i.e, 1,000 FDA approved drugs, 261 positive set drugs and 37 drugs from a study using machine learning (**Suvarna et al., 2021**). The FDA approved drugs were extracted from the DrugBank database (https://go.drugbank.com/) used for input for the server (**Supplementary Table 6**). The intermediate result file has been analysed and found that desmopressin, aspartic acid, alitretinoin, sirolimus and ibuprofen were top five FDA approved drugs with maximum targets found in the TTD database (**Supplementary Table 7**). The distribution of the FDA approved drugs is shown in **Supplementary Figure 2**.

The positive drugs dataset (261 drugs) was collected from clinical reports and publications in literature. Those drugs were subjected to analysis with this tool and the intermediate file was obtained (**Supplementary Table 8**). Curcumin, ellagic acid, p-coumaric acid, sirolimus and ibuprofen were the top five drugs found after the analysis of the intermediate file (**Supplementary Table 9**). We have plotted the distribution of the total number of common targets against the total number of drugs (**Supplementary Figure 3**). Total 157 drug's targets were found common in the both datasets and the number of common targets is 1.

In another experiment, we have extracted 37 drugs from the study reported by Suvarna et al in the year 2021 (**Suvarna et al., 2021**). Suverna et al predicted 37 drugs as the prognostic markers for the COVID-19 using proteomics and machine learning approaches. Those drugs were used as samples for our server and the predicted intermediate file was collected (**Supplementary Table 10**). The resultant file was analysed and top drugs were found including epigallocatechin gallate,

ponatinib, rapamycin, valproic acid and ruxolitinib which consist of maximum targets (**Supplementary Table 11**). Further, we have plotted the distribution of the drugs with the common drugs found (**Supplementary Figure 4**).

The intermediate file of results consists of 12 columns. The "ID" column represents the serial number of the drugs starting from 0, "DRUG" column represents the drug name, the "IN_DATABASE_VALUE" column describes the availability of the drug (ranges from 1 to -1), the "VALUE" column gives the module prediction (either 0 or 1). The "TOTAL_NUMBER_OF_COVID19_TARGETS", "NUMBER_OF_TARGETS", and "NUMBER_OF_COMMON_TARGETS" columns show the total number of targets found from the COVID-19 dataset, the total number of targets found from the TTD dataset and the total number of common targets found in both datasets, respectively. The "SIMILARITY" column calculates the similarity of the target on the basis of total targets found in the TTD dataset. The "SIMILARITY_COVID19" column calculates the similarity of the targets on the basis of Covid-19 targets. The "PERCENT_SIMILARITY" column predicts the percentage probability of occurrence of a target in the dataset. The "COMMON_TARGETS" and "ALL_TARGETS" columns provide the name of common targets and all drug-associated targets listed in the dataset obtained from the TTD database.

The user can observe N/A or 0 value in some columns which means the given drugs are not available in the database will not analyse and predict N/A in NUMBER_OF_TARGETS_TTD, NUMBER_OF_COMMON_TARGETS and PERCENT_SIMILARITY columns and 0 value in IN_DATABASE_VALUE column.

The formula to calculate similarity is as follows:

$$\text{Similarity} = \frac{\text{Common targets}}{\text{Total targets in TTD}}$$

The formula to calculate similarity is as follows:

$$\text{Similarity\_Covid19} = \frac{\text{Common Targets}}{\text{Covid-19 Targets}}$$

The formula to calculate percentage similarity is as follows:

$$\text{Percentage similarity} = \frac{\text{Total number of common targets}}{\text{Total number of targets}} \times 100$$

If, for example, the common target = 1, and the total target = 2, then the Percentage similarity = 50%.

When we screened the list of molecules from databases, we found that curcumin shows maximum overlap amongst its target space. For example, out of 19 curcumin targets, 7 targets are present in the pool of COVID-19 drug targets. The common targets are Carbonic Anhydrase Iv (Ca-Iv), Carbonic Anhydrase I (Ca-I), Carbonic Anhydrase Ii (Ca-Ii), Prostaglandin G/H Synthase 2 (Cox-2), Prostaglandin G/H Synthase 1 (Cox-1), Xanthine Dehydrogenase/Oxidase (Xdh) And Carbonic Anhydrase Ix (Ca-Ix).

5. Conclusion

Drug-target interaction prediction is an important part of most of the rational drug repositioning approaches. In fact, different biochemical, physical, and mathematical techniques have been designed and optimized to accurately infer links between ligands and proteins.
In this work, we utilize the drug target database (TTD database) for the prediction of off-target effects to suggest potential cases of drug repurposing and determine the molecular mechanism responsible for side-effects. There are a number of ways to find out the drug-off targets. Here we have combined the experimental data and system biology approach to yield a promising tool to better understand the biological response of the drug.

6. References

1. Chen S, Jiang H, Cao Y, Wang Y, Hu Z, Zhu Z, Chai Y. Drug target identification using network analysis: taking active components in Sini decoction as an example. Scientific reports. 2016 Apr 20;6(1):1-4.
2. Parisi D, Adasme MF, Sveshnikova A, Bolz SN, Moreau Y, Schroeder M. Drug repositioning or target repositioning: A structural perspective of drug-target-indication relationship for available repurposed drugs. Computational and structural biotechnology journal. 2020 Jan 1;18:1043-55.
3. Wang Y, Li F, Zhang Y, Zhou Y, Tan Y, Chen Y, Zhu F. Databases for the targeted COVID‑19 therapeutics. British Journal of Pharmacology. 2020 Nov;177(21):4999-5001.
4. Guy, R. K., DiPaola, R. S., Romanelli, F., & Dutch, R. E. (2020). Rapid repurposing of drugs for COVID-19. *Science*, 368, 829–830

5. Zhang, H., Penninger, J. M., Li, Y., Zhong, N., & Slutsky, A. S. (2020). Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Medicine*, 46, 586–590

6. T Issa N, J Peters O, W Byers S, Dakshanamurthy S. RepurposeVS: a drug repurposing-focused computational method for accurate drug-target signature predictions. Combinatorial chemistry & high throughput screening. 2015 Sep 1;18(8):784-94.

7. Fox S, Farr-Jones S, Yund MA. High-throughput screening for drug discovery: continually transitioning into new technologies. *J Biomol Screen.* 1999;4:183–186.

8. Griffith M, Griffith O, Coffman AC, Weible JV, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM, Miller CA, Subramanian J, Govindan R, Kumar RD, Bose R, Ding L, Walker JR, Larson DE, Dooling DJ, Smith SM, Ley TJ, Mardis ER, Wilson RK. DGIdb: mining the druggable genome. *Nat Methods.* 2013;10:1209–1210.

9. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model.* 2012;52:1757–1768.

10. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein Pharmacology by ligand chemistry. *Nat Biotechnol.* 25:197–206.

11. Warner WA, Sanchez R, Dawoodian A, Li E, Momand J. Identification of FDA-approved drugs that computationally bind to MDM2. *Chem Biol Drug Des.* 2012;80:631–637.

12. Bolton EE, Chen J, Kim S, Han L, He S, Shi W, Simonyan V, Sun Y, Thiessen PA, Wang J, Yu B, Zhang J, Bryant SH. PubChem3D: a new resource for scientists. *J Cheminform.* 2011;3:32.

13. Chen YZ, Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecules. *Proteins.* 2001;43:217–226.

14. Kellenberger E, Foata N, Rognan D. Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. *J Chem Inf Model.* 2008;48:1014–1025.14.

15. Kellenberger E, Schalon C, Rognan D. How to measure the similarity between protein ligand-binding sites? *Curr Comput Aided Drug Des.* 2008;4:209–220.

16. Jagannadham J, Jaiswal HK, Agrawal S, Rawal K. Comprehensive Map of Molecules Implicated in Obesity. PLoS One. 2016 Feb 17;11(2):e0146759. doi: 10.1371/journal.pone.0146759. PMID: 26886906; PMCID: PMC4757102.

17. Rawal K, Sinha R, Abbasi BA, Chaudhary A, Nath SK, Kumari P, Preeti P, Saraf D, Singh S, Mishra K, Gupta P, Mishra A, Sharma T, Gupta S, Singh P, Sood S, Subramani P, Dubey AK, Strych U, Hotez PJ, Bottazzi ME. Identification of vaccine targets in pathogens and design of a vaccine using computational approaches. Sci Rep. 2021 Sep 2;11(1):17626. doi: 10.1038/s41598-021-96863-x. PMID: 34475453; PMCID: PMC8413327.

18. P, Preeti, Robin Sinha, and Kamal Rawal. 2021. "Mex Pipeline for Analysis of Mobile Genetic Elements in Cancer Genome." OSF Preprints. November 25. doi:10.31219/osf.io/7ywnm.

19. Zhou Y, Zhang Y, Lian X, Li F, Wang C, Zhu F, Qiu Y, Chen Y. Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. Nucleic Acids Res. 2021 Oct 28:gkab953. doi: 10.1093/nar/gkab953. Epub ahead of print. PMID: 34718717.

20. Suvarna K, Biswas D, Pai MG, Acharjee A, Bankar R, Palanivel V, Salkar A, Verma A, Mukherjee A, Choudhury M, Ghantasala S. Proteomics and Machine Learning Approaches Reveal a Set of Prognostic Markers for COVID-19 Severity With Drug Repurposing Potential. Frontiers in physiology. 2021 Apr 27;12:432.

COV-Drug Target interaction server for Covid-19 Drug Repurposing

Kamal Rawal[#1], Prashant Singh[1], Robin Sinha[1], Priya Kumari[1], Swarsat Kaushik Nath[1], Sukriti Sahai[1], Ridhima[1], Sweety Dattatraya Shinde[1], Nikita Garg[1], Preeti P.[1], Trapti Sharma[1]

1. Amity Institute of Biotechnology, Amity University, Uttar Pradesh, India

#Corresponding Author
Email ID: kamal.rawal@gmail.com
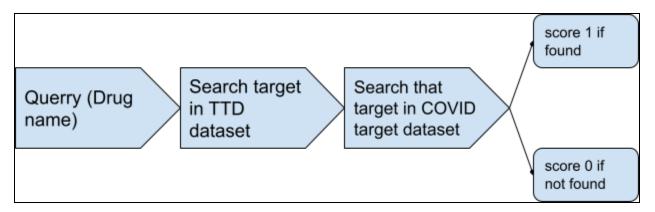Centre for Computational Biology and Bioinformatics, AIB
Amity University, Noida.

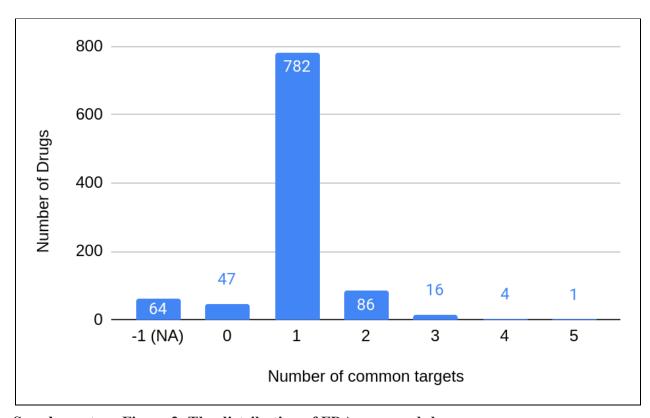**Keywords:** Bioinformatics, drug repurposing, artificial intelligence, COVID-19, molecular targets

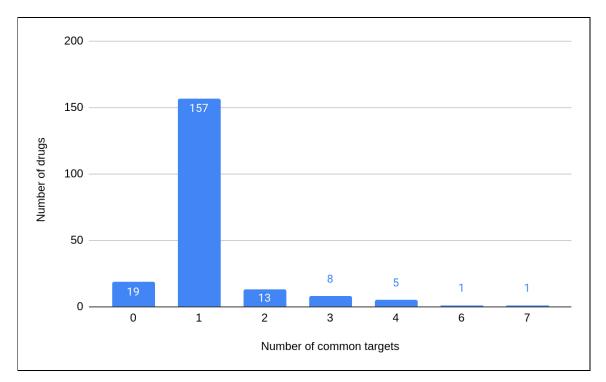**Supplementary Data Website**: https://sites.google.com/view/drugx-supplementary

COV-DRUGX Software Pipeline : http://drugx.kamalrawal.in/drugx/
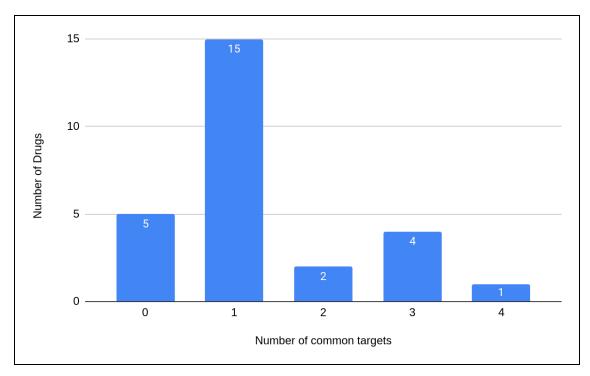
Supplementary Figure 1: The schematic diagram of the working of the drug target tool.

Supplementary Figure 2: The distribution of FDA approved drugs.

Supplementary Figure 3: The distribution of positive drugs dataset.

Supplementary Figure 4: The distribution of ML drugs dataset.