CNS Spectrums

www.cambridge.org/cns

Editorial

Cite this article: Kowshik AV, Manoj M, Sowmyanarayan S, and Chatterjee J (2024). Drug repurposing: databases and pipelines. *CNS Spectrums* **29**(1), 6–9.
https://doi.org/10.1017/S1092852923002365

Received: 27 June 2023
Accepted: 12 July 2023

Corresponding author:
Jhinuk Chatterjee;
Email: jhinukchatterjee@pes.edu

Drug repurposing: databases and pipelines

Arjun V. Kowshik, Megha Manoj, Siddarth Sowmyanarayan and Jhinuk Chatterjee

Department of Biotechnology, PES University, Bengaluru, India

Abstract

The concept of drug repurposing is focused on the repositioning of drug molecules that have already undergone safety trials. There are different strategies for drug repurposing. Network-based strategy focuses on the evaluation of drug combinations in a molecular environment with multi-target hits and analysis of drug interactions. Implementation of any in silico strategy requires several databases and pipelines for executing the process of shortlisting appropriate drugs.

Adherence to the notion of one drug used for one protein as treatment has failed to provide productive solutions for several disorders caused by multiple diverse factors. Currently, pharmaceutical companies are aiming to reposition approved drugs as treatment for various diseases as de novo drug designing is not only slow, but also cumbersome.

The concept of drug repurposing is focused on the sensible repositioning of a plethora of drug molecules that have already undergone safety trials thereby avoiding extensive investment. Drug repurposing is evaluated as a low-risk and highly efficient strategy due to the investment of $1.6 billion, which is nearly one-tenth of the monetary commitment required for de novo designing of a drug,[1] antibiotic resistance,[2] rare diseases,[3] oncological studies,[4] inflammatory disorders,[5] and neurological diseases[6] highlights its importance.

The scope of drug repurposing broadened massively since the COVID-19 pandemic and most of the drugs used to treat moderate to severe COVID-19 infections were initially screened using repurposing pipelines.[7]

Different approaches like signature matching, genetic association, and pathway mapping can be used for drug repurposing.[8] Two main strategies are namely drug-based, wherein a known FDA-approved drug is chosen as the starting point and either on-target or off-target drug repositioning is performed, and target-based, wherein a target molecule or biomarker is chosen as the starting and all FDA-approved drugs are screened against this target molecule using computational approaches.[9]

Network-based approach facilitates the evaluation of drug combinations in a molecular environment with multi-target hits and analysis of drug interactions.[10] One of the key strategies is interactome construction using different genes and related gene products.[11] Networks called diseasomes link multiple disorders and disease-related genes to provide a better insight into how different diseases are linked to each other and this helps pinpoint those genes which form a major link between seemingly unrelated diseases and behave as targets of drug repositioning.[11] Network medicine has proven to play a vital role in drug repurposing in identifying the drug targets by calculating the proximity scores between drug targets and disease genes in the human protein–protein interactome.[12]

Implementation of any in silico strategy requires extensive research using databases, and pipelines for executing the process of shortlisting appropriate drugs. Every drug repositioning pipeline requires a collection of all the FDA-approved drugs and their pharmacological properties, structural information of drugs and proteins, gene-level analytical resources, knowledge of the disease, and the pathways related to the disease.

There are different categories of databases commonly used in drug repurposing.

Molecular structure of drugs: ChEMBL[13] (https://www.ebi.ac.uk/chembl/) is a manually curated database containing drugs and molecules with high bioactivity. PubChem[14] (https://pubchem.ncbi.nlm.nih.gov/) is a database of chemical molecules by name, structure, molecular formula, biological activities, physical properties, safety, and other identifiers. DrugBank[15] (https://go.drugbank.com/) is a manually curated comprehensive database of drugs and their targets, groups, pharmacological parameters, interactions, transporters, mechanism of action, and useful properties. ChemSpider[16] (http://www.chemspider.com/) is a chemical structure database providing all the properties, literature review, vendors, and similar molecules to a particular ligand.

Drug targets: BindingDB[17] (https://www.bindingdb.org/rwd/bind/index.jsp) provides information on binding affinities and interactions of all ligand molecules with the target and off-target proteins. Drug Target Commons[18] (https://drugtargetcommons.fimm.fi/) provides annotations

regarding bioactivities of drug–target interactions and relation with multiple proteins in the body accompanied by bioassay data. STITCH[19] (http://stitch.embl.de) produces a detailed map of the drug–protein interactions and further protein–protein relations to identify all possible drug targets. TTD[20] (https://db.idrblab.net/ttd/) is an organized platform for identifying drugs for a specific target protein along with patient data and literature. DrugCentral[21] (https://drugcentral.org/) has up-to-date drug information based on protein targets for repositioning.

Drug side effects and toxicity: SIDER[22] (http://sideeffects.embl.de/) is a collection of the drug indications and recorded adverse drug reactions of all marketed drugs. TOXRIC[23] (https://toxric.bioinforai.tech/home) is a novel repository with metabolic reactions and categorized toxicological data for all drugs.

Protein and Gene related: UniProt[24] (https://www.uniprot.org/) has the highest quality information regarding the structural and functional annotation of discovered proteins. GproteinDB[25] (https://gproteindb.org/?) stores structural, functional, mutational, and endogenous ligand information of all G-coupled proteins in the body. PDB[26] (https://www.rcsb.org/) has 3-D X-ray crystallographic and NMR structures of proteins along with important structural annotations. SCOPe[27] (https://scop.berkeley.edu/) is a highly curated structural classification of proteins focusing on protein similarity and interactions. HuRI[28] (http://interactome-atlas.org/) has mapping of the interaction of the target protein with all other proteins in the body along with their relationships. HINT[29] (http://hint.yulab.org/) stores binary and co-complex protein interactions mapped to generate interactome. STRING[30] (https://string-db.org/) provides all the functional associations of a protein in the form of an interactome along with sequence similarity, disease pathway, and latest annotations. OMIM[31] (https://www.omim.org/) is a thorough assembly of all human genes with reference to phenotypic expression in disorders, structure, location, mapping, and functioning. GenBank[32] (https://www.ncbi.nlm.nih.gov/genbank/) is genetic sequence collection with protein translation information and gene similarities. BioGRID[33] (https://thebiogrid.org/) stores detailed description of the gene–gene and gene-chemical interactions. DGIdb[34] (https://www.dgidb.org/) has curated details on all the gene interactions for a drug and vice versa. IntAct[35] (https://www.ebi.ac.uk/intact/home) provides molecular interactions specifically protein–protein and RNA-protein interactions. GEO[36] (https://www.ncbi.nlm.nih.gov/geo/) accepts array and sequence-based data and provides tools to query and download experiments and curated gene expression profiles.

Pathways: KEGG[37] (https://www.genome.jp/kegg/) stores detailed map of the molecular reactions, relations, and interactions along with the biochemistry of the disease. Reactome[38] (https://reactome.org/) has representation of entire pathways, proteins, reactions, and interactions specific to diseases and protein complexes in a hierarchical manner. Pathway Commons[39] (https://www.pathwaycommons.org/) is used for illustrious pathway visualization for various biochemical processes and diseased states.

Disease-associated: Phenopedia[40] (https://phgkb.cdc.gov/PHGKB/startPagePhenoPedia.action) offers summarized information in a disease-centric view on human genetic associations. DisGeNET[41] (https://www.disgenet.org/) is a large collection of genes and variants associated with human diseases and integrates data from various repositories to provide various gene-disease and variant-disease associations. DISEASES—Database Commons[42] (https://diseases.jensenlab.org/) provides records of the major genes, proteins, and experiments relating to a disease.

Repurposed Drugs: ReDO_DB[43] (https://www.anticancerfund.org/en/redo-db) stores information on compounds that are non-cancer drugs showing anticancer activity. DrugRepV[44] (https://bioinfo.imtech.res.in/manojk/drugrepv/) stores manually curated drugs and chemicals that display antiviral activity against epidemic and pandemic viruses. ReFRAME[45] (https://reframedb.org/) contains nearly all small molecules that have reached clinical development or undergone significant preclinical profiling.

Drug repurposing pipelines focus on obtaining the best set of candidate drug molecules for the selected disease through rigorous computational techniques and intensified analytical measures.

Computational Analysis of Novel Drug Opportunities (CANDO)[46] (https://github.com/ram-compbio/CANDO) is an open-source platform for the analysis of drug interactions on a proteomic scale. In CANDO drug–protein interaction signatures are generated from an extensive library of known interaction mappings and compared and screened based on their similarities to the drugs used for the same disease. In total, 51 of the 276 molecules predicted by this platform against SARS-CoV-2 are explored in clinical studies and have demonstrated promising activity.[47]

KsRepo[48] (https://github.com/adam-sam-brown/ksRepo) is an R-based open-source expression-level platform for drug repurposing that identifies potential candidates enrichment scores. KsRepo analyses and compares RNA-seq data to known gene–drug interactions and also exhibits flexibility in data set types and can also predict drug candidates with limited information on drug–gene interactions.[49]

SperoPredictor[50] is a generic repurposing framework consisting of multiple machine learning algorithms like Random Forest, Tree Ensemble, and Gradient Boosted Trees to predict repurposable drug candidates. SperoPredictor pipeline involves data collection, training, and subsequent deployment of machine learning models followed by literature and molecular docking-based validation.

Single-cell Guided Pipeline to Aid Repurposing of Drugs (ASGARD)[51] (https://github.com/lanagarmire/ASGARD) makes use of scRNA-seq data of single cell and cell clusters in the disease to predict potentially repurposable drugs. ASGARD processes scRNA-seq data for differential gene analysis and identifies drug candidates which reverse gene expression.

One of the most recent approaches to drug repurposing is using network medicine to identify biomarkers and target molecules through interactomes.[52] Searching off-lAbel dRUg aNd NEtwoRk (SAveRUNNER)[53] (https://github.com/sportingCode/SAveRUNNER) is a R-based tool that predicts drugs that can be repositioned by preparing a drug-disease association with the help of human interactome. SAveRUNNER prepares a drug-disease association by calculating the network-based proximity scores which show how close the drug molecule is to the disease in the human interactome. It has been successfully used to repurpose drugs for diseases such as COVID-19,[54] Amyotrophic Lateral Sclerosis,[55] Breast cancer,[56] and cardiovascular disease.[57]

The Konstanz-Integration Miner (KNIME)[58] (www.knime.com) is an open-source platform used to generate semi-automated workflows which are intensely used for drug repurposing. The approach used in KNIME is based on mining of data from various exhaustive databases and processing, analyzing, and visualizing this data in order to shortlist the drugs that can be repurposed.

https://doi.org/10.1017/S1092852923002365 Published online by Cambridge University Press

Conclusion

With an increasing demand to find cures for diseases using faster and more cost-effective methods, drug repurposing is proving to be the future of medicine and the more sensible approach as compared to de novo drug design by reducing the number of clinical and experimental trials and making strong predictions about which drugs can be repositioned, thus, paving the way for successful treatments.

Financial support. There was no funding source pertaining to this research work.

Author contribution. Conceptualization: J.C.; Formal analysis: A.V.K., M.M., S.S.; Methodology: A.V.K., M.M., S.S.; Project administration: J.C.; Resources: A.V.K., M.M., S.S.; Software: A.V.K., M.M., S.S.; Supervision: J.C.; Writing—original draft: A.V.K., M.M., S.S.; Writing—review and editing: J.C.

Disclosure. The authors do not have any competing interests to disclose.

References

1. Xue H, Li J, Xie H, Wang Y. Review of drug repositioning approaches and resources. *Int J Biol Sci.* 2018;**14**(10):1232–1244.
2. Younis W, Thangamani S, Seleem MN. Repurposing non-antimicrobial drugs and clinical molecules to treat bacterial infections. *Curr Pharm Des.* 2015;**21**(28):4106–4111.
3. Roessler HI, Knoers NVAM, van Haelst MM, van Haaften G. Drug repurposing for rare diseases. *Trends Pharmacol Sci.* 2021;**42**(4):255–267.
4. Schein CH. Repurposing approved drugs for cancer therapy. *Br Med Bull.* 2021;**137**(1):13–27.
5. Kingsmore KM, Grammer AC, Lipsky PE. Drug repurposing to improve treatment of rheumatic autoimmune inflammatory diseases. *Nat Rev Rheumatol.* 2020;**16**(1):32–52.
6. Massey TH, Robertson NP. Repurposing drugs to treat neurological diseases. *J Neurol.* 2018;**265**(2):446–448.
7. Galindez G, Matschinske J, Rose TD, et al. Lessons from the COVID-19 pandemic for advancing computational drug repurposing strategies. *Nat Comput Sci.* 2021;**1**(1):33–41.
8. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov.* 2019;**18**(1):41–58.
9. Spitschak A, Gupta S, Singh KP, Logotheti S, Pützer BM. Drug repurposing at the interface of melanoma immunotherapy and autoimmune disease. *Pharmaceutics.* 2022;**15**(1):83.
10. Wu Z, Wang Y, Chen L. Network-based drug repositioning. *Mol Biosyst.* 2013;**9**(6):1268–1281.
11. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011;**12**(1):56–68.
12. Cheng F, Desai RJ, Handy DE, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun.* 2018;**9**:2691.
13. Davies M, Nowotka M, Papadatos G, et al. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* 2015;**43**(W1):W612–W620.
14. Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res.* 2016;**44**(D1):D1202–D1213.
15. Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008;**36**(Database issue):D901–D906.
16. Pence HE, Williams A. ChemSpider: an online chemical information resource. *J Chem Educ.* 2010;**87**(11):1123–11242010.
17. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 2007;**35**(Database issue):D198–D201.
18. Tang J, Tanoli ZU, Ravikumar B, et al. Drug target commons: a community effort to build a consensus knowledge base for drug-target interactions. *Cell Chem Biol.* 2018;**25**(2):224–229.e2.
19. Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.* 2008;**36**(Database issue):D684–D688.
20. Zhou Y, Zhang Y, Lian X, et al. Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res.* 2022;**50**(D1):D1398–D1407.
21. Avram S, Bologa CG, Holmes J, et al. DrugCentral 2021 supports drug discovery and repositioning. *Nucleic Acids Res.* 2021;**49**(D1):D1160–D1169.
22. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* 2016;**44**(D1):D1075–D1079.
23. Wu L, Yan B, Han J, et al. TOXRIC: a comprehensive database of toxicological data and benchmarks. *Nucleic Acids Res.* 2023;**51**(D1):D1432–D1445.
24. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015;**43**(D1):D204–D212.
25. Pándy-Szekeres G, Esguerra M, Hauser AS, et al. The G protein database, GproteinDb. *Nucleic Acids Res.* 2022;**50**(D1):D518–D525.
26. Berman HM, Westbrook J, Feng Z, et al. The protein data bank. *Nucleic Acids Res.* 2000;**28**(1):235–242.
27. Chandonia JM, Guan L, Lin S, Yu C, Fox NK, Brenner SE. SCOPe: improvements to the structural classification of proteins – extended database to facilitate variant interpretation and machine learning. *Nucleic Acids Res.* 2022;**50**(D1):D553–D559.
28. Luck K, Kim DK, Lambourne L, et al. A reference map of the human binary protein interactome. *Nature.* 2020;**580**(7803):402–408.
29. Das J, Hint YH. High-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol.* 2012;**6**(1):1–12.
30. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;**49**(D1):D605–D612.
31. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 2015;**43**(Database issue):D789–D798.
32. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res.* 2005;**33**(Database issue):D34–D38.
33. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 2006;**34**(Database issue):D535–D539.
34. Cotto KC, Wagner AH, Feng YY, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 2018;**46**(D1):D1068–D1073.
35. Hermjakob H, Montecchi-Palazzi L, Lewington C, et al. IntAct: an open source molecular interaction database. *Nucleic Acids Res.* 2004;**32**(Database issue):D452–D455.
36. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002;**30**(1):207–210.
37. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;**28**(1):27–30.
38. Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* 2018;**46**(D1):D649–D655.
39. Cerami EG, Gross BE, Demir E, et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* 2011;**39**(Database issue):D685–D690.
40. Yu W, Clyne M, Khoury MJ, Gwinn M. Phenopedia and Genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. *Bioinformatics.* 2010;**26**(1):145–146.
41. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 2020;**48**(D1):D845–D855.
42. Grissa D, Junge A, Oprea TI, Jensen LJ. Diseases 2.0: a weekly updated database of disease-gene associations from text mining and data integration. *Database (Oxford).* 2022;**2022**:baac019.
43. Pantziarka P, Verbaanderd C, Sukhatme V, et al. ReDO_DB: the repurposing drugs in oncology database. *Ecancermedicalscience.* 2018;**12**:886.

https://doi.org/10.1017/S1092852923002365 Published online by Cambridge University Press

44. Rajput A, Kumar A, Megha K, Thakur A, Kumar M. DrugRepV: a compendium of repurposed drugs and chemicals targeting epidemic and pandemic viruses. *Brief Bioinform.* 2021;**22**(2):1076–1084.
45. Janes J, Young ME, Chen E, et al. The ReFRAME library as a comprehensive drug repurposing library and its application to the treatment of cryptosporidiosis. *Proc Natl Acad Sci U S A*. 2018;**115**(42):10750–10755.
46. Mangione W, Falls Z, Chopra G, Samudrala R. cando.py: open source software for predictive bioanalytics of large scale drug-protein-disease data. *J Chem Inf Model.* 2020;**60**(9):4131–4136.
47. Mangione W, Falls Z, Samudrala R. Optimal COVID-19 therapeutic candidate discovery using the CANDO platform. *Front Pharmacol.* 2022;**13**: 970494.
48. Brown AS, Kong SW, Kohane IS, et al. ksRepo: a generalized platform for computational drug repositioning. *BMC Bioinform*. 2016;**17**:78.
49. Traylor JI, Sheppard HE, Ravikumar V, et al. Computational drug repositioning identifies potentially active therapies for chordoma. *Neurosurgery.* 2021;**88**(2):428–436.
50. Ahmed F, Lee JW, Samantasinghar A, et al. SperoPredictor: an integrated machine learning and molecular docking-based drug repurposing framework with use case of COVID-19. *Front Public Health*. 2022; **10**:902123.
51. He B, Xiao Y, Liang H, et al. ASGARD is a single-cell guided pipeline to aid repurposing of drugs. *Nat Commun.* 2023;**14**:993.
52. Caldera M, Buphamalai P, Müller F, Menche J. Interactome-based approaches to human disease. *Curr. Opin. Syst. Biol.* 2017;**3**:88–94.
53. Fiscon G, Paci P. SAveRUNNER: an R-based tool for drug repurposing. *BMC Bioinform*. 2021;**22**(1):150.
54. Fiscon G, Conte F, Farina L, Paci P. SAveRUNNER: a network-based algorithm for drug repurposing and its application to COVID-19. *PLoS Comput Biol.* 2021;**17**(2):e1008686.
55. Fiscon G, Conte F, Amadio S, Volonté C, Paci P. Drug repurposing: a network-based approach to amyotrophic lateral sclerosis. *Neurotherapeutics.* 2021;**18**(3):1678–1691.
56. Conte F, Sibilio P, Fiscon G, Paci P. A transcriptome- and interactome-based analysis identifies repurposable drugs for human breast cancer subtypes. *Symmetry.* 2022;**14**(11):2230.
57. Paci P, Fiscon G, Conte F, et al. Comprehensive network medicine-based drug repositioning via integration of therapeutic efficacy and side effects. *NPJ Syst Biol Appl.* 2022;**8**(1):12.
58. Tuerkova A, Zdrazil B. A ligand-based computational drug repurposing pipeline using KNIME and Programmatic Data Access: case studies for rare diseases and COVID-19. *J Cheminform.* 2020;**12**:71.

https://doi.org/10.1017/S1092852923002365 Published online by Cambridge University Press