

REVIEW ARTICLE OPEN



Strategies to identify candidate repurposable drugs: COVID-19 treatment as a case example

Ali S. Imami ¹, Robert E. McCullumsmith^{1,2} and Sinead M. O'Donovan ¹✉

© The Author(s) 2021

Drug repurposing is an invaluable strategy to identify new uses for existing drug therapies that overcome many of the time and financial costs associated with novel drug development. The COVID-19 pandemic has driven an unprecedented surge in the development and use of bioinformatic tools to identify candidate repurposable drugs. Using COVID-19 as a case study, we discuss examples of machine-learning and signature-based approaches that have been adapted to rapidly identify candidate drugs. The Library of Integrated Network-based Signatures (LINCS) and Connectivity Map (CMap) are commonly used repositories and have the advantage of being amenable to use by scientists with limited bioinformatic training. Next, we discuss how these recent advances in bioinformatic drug repurposing approaches might be adapted to identify repurposable drugs for CNS disorders. As the development of novel therapies that successfully target the cause of neuropsychiatric and neurological disorders has stalled, there is a pressing need for innovative strategies to treat these complex brain disorders. Bioinformatic approaches to identify repurposable drugs provide an exciting avenue of research that offer promise for improved treatments for CNS disorders.

Translational Psychiatry (2021)11:591 ; <https://doi.org/10.1038/s41398-021-01724-w>

INTRODUCTION

The outbreak of the COVID-19 pandemic led to an unprecedented response in the scientific community to rapidly identify, develop and implement pharmacotherapies for the treatment of COVID-19 and its causative agent, the novel *coronaviridae* strain severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In less than two years, over 6 000 clinical trials related to COVID-19 have been registered (clinicaltrials.gov) and almost 1 500 articles specifically related to drug repurposing for COVID-19 treatment can be found on PubMed (Fig. 1). Despite the massive investment of resources and a focused global effort, only a single drug, the broad spectrum antiviral remdesivir, is currently approved by the FDA for the treatment of COVID-19 [1]. This highlights the significant challenges to identifying pharmacotherapies for complex, heterogeneous diseases and is a familiar scenario to researchers in the field of neuropsychopharmacology. Although the initial urgency to identify drug treatments has waned with the development of effective vaccines, the emergence of new SARS-CoV-2 variants “of concern,” geographical inequities in vaccine availability and the predicted need for periodic booster shots ensure that effective drug treatments for COVID-19 are still an important tool to overcoming the pandemic. Here, we evaluate *in silico* approaches deployed to identify candidate repurposable drugs for the treatment of COVID-19 to determine what lessons can be gleaned from the accelerated and concerted efforts during the pandemic, and whether they can inform shortcomings in the identification of pharmacotherapies for CNS disorders. Using COVID-19 as a case study we will (1) summarize *in silico* methodologies applied to identify repurposable drugs (Fig. 2) and (2) discuss how they might inform current drug repurposing efforts for neuropsychiatric

disorders. Selected representative studies will be discussed that highlight how different drug discovery approaches may be adapted by neuroscience researchers to address the pressing need for improved treatments for psychiatric conditions.

COVID-19 CASE STUDY

The focus of drug discovery efforts for COVID-19 has rested on drug repurposing or repositioning, that is, identifying novel uses for existing (approved and investigational) drugs that are different from their original indication [2]. The advantages of this approach include significant savings in terms of time and cost. Drug repurposing circumvents some of the safety, drug tolerability and dosage concerns associated with developing a drug lead from bench to bedside, shaving years off the drug development pipeline [3].

Within weeks to months of the COVID-19 pandemic, genomic, transcriptomic, and proteomic profiles were available for the causative SARS-CoV-2 virus and from infected human tissues. Data sharing was prioritized by institutions around the world and a surge of research papers soon followed. *In silico* approaches were deployed to mine the available data without the need for laboratory wet work in an era of quarantine and remote working. Established bioinformatic workflows were quickly adapted and applied to identify therapeutic targets and candidate repurposable drugs for the treatment of COVID-19.

MACHINE LEARNING APPROACHES

The advancement in computer science and data storage have brought forth what could be easily termed as the “Age of Artificial

¹Department of Neurosciences, University of Toledo, Toledo, OH, USA. ²Neurosciences Institute, Promedica, Toledo, OH, USA. ✉email: sinead.odonovan@utoledo.edu

Received: 7 September 2021 Revised: 26 October 2021 Accepted: 2 November 2021

Published online: 16 November 2021

Intelligence (AI). “Deep Learning,” one of the methods to achieve AI in a particular system, has become the de-facto standard and is often used synonymously with AI. The field of biomedical research is not an exception to this trend. Over the past decade, numerous attempts have been made to integrate AI systems into diagnostics, epidemiology, outcome prediction and drug discovery [4–6]. There has been mixed levels of success in these fields and there is still a long way to go before AI can be used autonomously for drug discovery. However, at present, Deep Learning and AI allows for the rapid synthesis of heterogeneous data sources to narrow down lists of potential compounds of therapeutic value [4–8]. AI-based approaches have three major components, all subject to their own constraints and limitations. First and foremost is data availability. Second is the processing of the data sources and selection and application of a particular AI

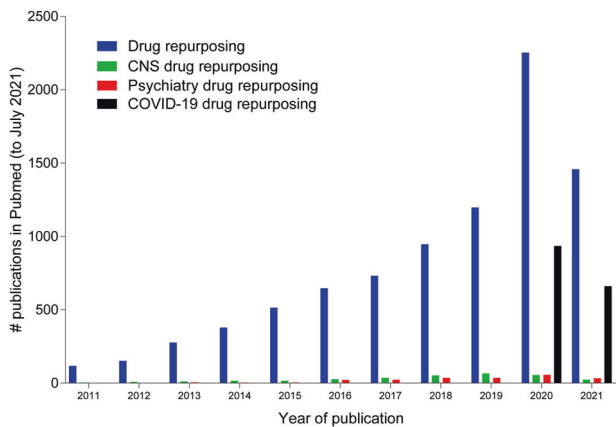


Fig. 1 The number of publications referencing “drug repurposing” has increased year-on-year over the last decade. Psychiatry and CNS drug repurposing reflect a small proportion of total publications. By comparison, in under two years, COVID-19 related drug repurposing efforts have resulted in the generation of a large number of publications.

technique. Finally, there is the dual problem of results interpretation and predictive validity.

Data Sources

Any machine learning (ML) approach requires high quality input data. This data will not only underlie the inferences that the AI/ML approach will generate but will also dictate the nature of the approach that is used [9]. Thus, it is an absolute necessity to generate databases that are easily accessible. Currently, there are several databases available online that encompass a variety of “omics.” These include primary databases like the National Center for Biotechnology Information’s (NCBI), Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>), PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) and the European Bioinformatics Institute’s ChEMBL (<https://www.ebi.ac.uk/chembl/>) as well as secondary databases such as ExCAPE-DB [10] and PhIN [11] that curate data from literature and primary databases creating an inferential layer of data [12]. In addition, there are other data sources that generate predictions based on existing datasets. These can include predicted protein-protein [13] and drug-receptor [11] interactions. Finally, other resources include databases like DrugR + [14] where empirically defined and predicted interactions between drugs and their possible replacements (candidate repurposable drugs) has been curated. In the wake of the SARS-CoV-2 pandemic, several of these data sources were utilized to identify novel drugs and to repurpose existing drugs. Chief among these were the Library of Interconnected Network-based Cellular Signatures (LINCS) [15–17], DrugBank [18–20] and Protein Databank [21–23].

Data processing

Once high-quality data sources have been identified, the next step is to integrate and process the data. For this purpose, there are several approaches that can be relied upon, including Logistic Regression, Support Vector Machines, and Convolutional and Recurrent Neural Networks [9]. Selection of a particular model is guided by the availability of appropriate datasets, the nature of the data itself and the expertise of the team conducting the

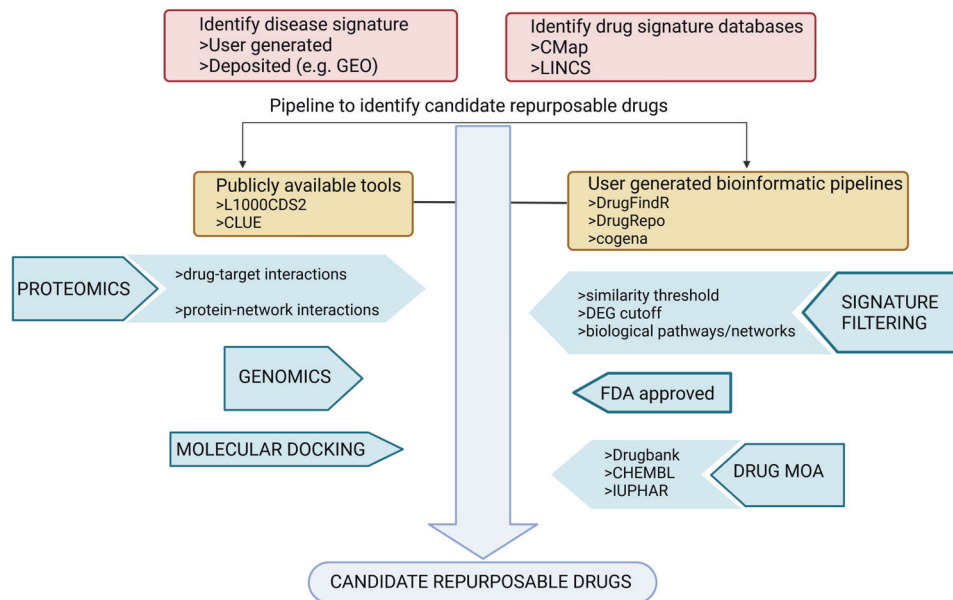


Fig. 2 Bioinformatic drug repurposing workflows. Following selection of transcriptomic disease data and drug signature databases, a large number of publicly available or user-designed tools can be adapted to identify candidate repurpose drugs. Signatures are generated based on different criteria, including differential gene expression (DEG) cutoff and disease x candidate drug similarity score. Genomic and proteomic data can also be integrated into the workflow to provide insight into drug-target interactions, for example. Candidate repurposable drugs may be further filtered by relevance of drug mechanism of action (MOA), using molecular docking to determine how a drug interacts with a target protein, or to those drugs that are already FDA-approved. Created with BioRender.com.

analysis. In particular, the Convolutional Neural Networks have become the de-facto standard for deep learning in recent years due to their ability to learn necessary parameters from the data autonomously. The ML models rely on specific parameters that are unique to each combination of dataset, method, and problem. These parameters are usually only significant in a mathematical sense and not interpretable by humans in a meaningful way. In contrast, “hyperparameter tuning” selects the values for the parameters that influence how the learning algorithm works. It is a time-consuming step in the process of building ML/AI based models, but it is essential to achieve optimal setup of these “hyperparameters” to generate models that can provide more accurate results without sacrificing time or computational efficiency. It is evident that AI/ML has matured significantly over the past decade [24, 25]. With projects like Virtual Physiological Human [26, 27] on the horizon, which aim to provide a functioning in silico simulation of a human that can react to stimuli, we may look forward to the transition of more in silico identified drugs to clinical trials soon [28].

For identifying SARS-CoV-2 treatments, deep learning approaches that relied on convolutional neural networks and that integrated a varied type of information, including transcriptomics, drug structure and protein sequences, appeared to work the best [29]. Other approaches relied on molecular docking simulations in addition to the above-mentioned datasets [30, 31]. As early as February 2020, machine learning and AI driven approaches showed promise in identifying candidate repurposable drugs for COVID-19. Using BenevolentAI, the numb-associated kinase (NAK) family was identified as a drug target for the treatment of COVID-19. The NAK inhibitor Baricitinib has a high affinity for AP2 associated protein kinase 1 (AAK-1) and was identified as the most promising candidate drug from this AI screen. Inhibition of AAK-1, a regulator of clathrin-mediated endocytosis, was expected to inhibit SARS-CoV-2 infection of host cells [32, 33]. Clinical trial (NCT04401579) of combination therapy remdesivir and baricitinib found fewer side effects and greater efficacy in reducing COVID-19 patient recovery time than with remdesivir treatment alone [34]. The corticosteroid dexamethasone was also identified by different groups using AI approaches, like that applied by AI VIVO and [35], resulting in identification of a therapy that reduced mortality in ventilated COVID-19 patients in clinical trial [36]. These studies highlight the utility of in silico and AI/ML based approaches to successfully identify high yield repurposable drugs.

SIGNATURE-BASED APPROACHES

A gene signature describes the unique pattern of gene expression in a cell associated with biological or pathological (disease) processes or in response to genetic or pharmacological perturbation. Highly similar signatures might then represent novel connections between a drug and protein target, or similar function of two structurally dissimilar chemical perturbagens. Conversely, if drug and disease signatures are dissimilar, then that drug may act to reverse the disease phenotype (signature reversion principle) [37]. Disease-related transcriptomic data is most often user-generated and made available through public repositories like NCBI GEO (<https://www.ncbi.nlm.nih.gov/geo/>). The need for a systematically generated catalog of cellular signatures led to the creation of LINCS [38], a product of a multi-institutional consortium. The LINCS platform (LINCS data portal 3.0) currently contains data on over 30,000 small molecules (<http://lincsportal.ccs.miami.edu/SmallMolecules/>) and over 2 million different chemical perturbagen transcriptomic signatures. It also provides a large range of tools for exploration and analysis of data. The L1000 is an assay of 978 landmark genes that provides a reduced representation (approximately 82%) of the transcriptome [39] and facilitated the systematic generation of gene signatures for almost 20,000 small molecules. The catalog of signature

connections is the connectivity map (CMap) [39, 40]. Tools for analysis of the massive L1000 datasets include the L1000 characteristic direction signature search engine (L1000CDS2) [41], iLINCS, a web platform for analysis of LINCS transcript and proteomic data and signatures [38], the Connectivity map (CMap), and LINCS Unified Environment (CLUE; <https://clue.io>). CLUE is a cloud-based query infrastructure that allows researchers to explore and seamlessly incorporate CMap data into their analyses. In addition, LINCS signature data is easily downloaded and integrated into user-designed pipelines to fit their experimental needs. In the hands of individual users, the signature data provided by the LINCS resources has generated hundreds of candidate repurposable drugs for the treatment of COVID-19 [17, 42–44]. An advantage of the LINCS and CMap platforms is their accessibility. With little or no formal bioinformatic training, the provided tools can be utilized by bench scientists to explore datasets of interest. With basic R skills, user-generated pipelines can be designed to interrogate LINCS datasets, as discussed below.

Examples of signature-based approaches to identify COVID-19 candidate repurposable drugs

Although almost all COVID-19 signature-based analyses use the same or similar publicly available RNAseq data from SARS-CoV-2 infected cell-lines and infected host tissues and cells [45, 46], different bioinformatic pipelines have identified hundreds of candidate drugs, perhaps raising concerns about the specificity and utility of such approaches. However, although hundreds of unique candidate drugs have been proposed for repurposing, these drugs are often members of the same drug class. Unsurprisingly, antiviral drugs were often reported as top hits, however, other antimicrobials, immunosuppressants and kinase inhibitors were also commonly reported [35, 43, 47]. Kinase inhibitors were of growing interest as antivirals prior to COVID-19 [48, 49]. The identification of drugs with immunosuppressant properties as top candidates [17, 50], amid the emergence of the “cytokine storm” as a major treatment concern during COVID-19, highlights the value of the signature-based approach. Additionally, signature-based approaches can be adapted for more targeted search efforts, for example in the case of COVID-19, to find drugs that act on the ACE2 receptor [51].

Here, we describe how signature-based approaches, in particular using the LINCS signature database, can be adapted to identify candidate drugs using publicly available tools or user generated bioinformatic pipelines (and could be applied to other disease states). Islam et al identified a set of hub genes from gene enrichment of SARS-CoV-2 transcriptome data and analysis of protein-protein interaction networks in lung tissue, for drug repositioning using LINCS [42]. Hub genes from the protein-protein interaction networks, generated using DifferentialNet Database [52] and NetworkAnalyst [52] tool and with a degree filter (>30), were considered to represent essential biological signaling molecules. The top 10 hub proteins were used for drug repurposing. Using the L1000CDS2 search engine, the chemical perturbagens with the lowest cosine similarity score for the input targets were identified. The top 10 chemical perturbagens were carried forward for further consideration; drug mechanism of action was checked in publicly available databases (e.g., Drugbank, IUPHAR) and the literature to ensure specificity. Of the top 10, kinase inhibitors and antimicrobials were represented, with 7 drugs having approval status and the remaining 3 having investigational status. Molecular docking analyses was carried out to further validate the bioinformatically determined candidate drugs.

Others develop bioinformatic pipelines to identify and curate candidate drugs on specific criteria deemed important for drug consideration. For example, using the R shiny application DrugFindR, our group extracted L1000 genes from publicly

available COVID-19 transcriptome datasets. Candidate repurposable drugs were identified if 1) they had chemical perturbagen signatures in at least 5/7 common cell lines in LINCS, 2) if the chemical perturbagen signature had a minimum discordance score of <-0.321 with COVID-19 disease signatures and 3) a concordance score >0.321 with known antivirals in use for COVID-19 treatment at that time and 4) if candidate drugs were FDA-approved [17]. Following manual curation of the literature and candidate drug list, a final “top hit” list was generated containing drugs that were already undergoing clinical trial for COVID-19, confirmed to specifically inhibit SARS-CoV-2 in vitro or had known *coronavirus* antiviral effects. Using a modified DrugFindR workflow, the potential mechanisms of action of two candidate drugs for the treatment of COVID-19, oxytocin [53] and fluoxetine [54] were also explored. Analysis of LINCS consensus gene knockdown signatures found that the oxytocin gene signature is highly concordant with inflammation knockdown signatures (e.g., IL-6 knockdown) but discordant with signatures of pro-immune marker knockdown, like CD40. Signature analysis also provided support for an anti-inflammatory role for the antidepressant fluoxetine, via the NF-kappaB-IL6-ST signal transduction pathway. Fluoxetine (NCT04377308) is currently undergoing clinical trial for COVID-19 and is expected to treat the cytokine storm associated with disease.

Alternative LINCS signature-based workflows include the DrugRepo pipeline (https://nelhachem.shinyapps.io/DrugRepo/?_ga=2.78354771.679352558.1592320962-118407775.1587897066) and [55]. This pipeline integrated LINCS chemical perturbagen gene signature data with gene and drug set enrichment analysis and drug-target associations to identify candidate repurposable drugs [55]. The authors curated drug-target interactions from DrugBank [18, 19], IUPHAR [56] and CLUE (<https://clue.io>). Drug set enrichment, which uses drug-target associations instead of gene sets, enriched for drug-target interactions among the identified candidate drugs. As an additional filter, protein targets with fewer than 3 representative drugs were removed. Importantly, analyzing a ruxolitinib-treated A549 cell line infected with SARS-CoV-2 (SARS2_A549_ACE2_RUXO; [45]) using this pipeline resulted in identification of ruxolitinib as a top hit. Drug-set enrichment analysis then confirmed that drugs of the ruxolitinib drug class, JAK2 inhibitors, were also overrepresented in their results, validating their approach and confirming the accuracy of the L1000 drug identification.

The co-expressed gene set enrichment analysis (cogena) is a framework that uses the CMap drug induced gene signatures to identify drugs that reverse differentially expressed and co-regulated genes [57]. Coexpressed genes are often expressed in the same biological pathways and are hypothesized as more likely drivers of the underlying disease biology, adding an extra dimension to selecting gene sets for targeting drug repositioning. This approach identified several drugs including two approved antiviral drugs (saquinavir and ribavirin) that were previously implicated by molecular docking approaches for the treatment of COVID-19 [44].

Signature-based analyses are most often based on gene expression (transcriptome) data, however, integrating protein-network data can add another dimension to analyses. Zhou et al devised a drug-target network of human coronavirus (HCoV)-host protein interactions. A network of 119 proteins that were targets of HCoV (6 different strains) or in pathways essential for infection by human coronaviruses were identified [58]. The drug-target network was compiled from drug-target interaction data from multiple sources including DrugBank [18, 19] and ChEMBL [59]. Drug-target interaction criteria included binding affinity ($EC_{50} \leq 10 \mu\text{M}$) and inclusion of a unique, reviewed accession number for that protein in UniProt. Druggable proteins were then mapped into the HCoV-host interactome where the authors found almost 40% of the protein interactome was targeted by at least one

(approved or under clinical trial) drug. To capitalize on the benefits of a network-based targeting approach [60–63], the human protein-protein interactome was expanded and a network proximity measure was applied to quantify connections in the HCoV-specific subnetwork, resulting in a larger list of 135 candidate repurposable drugs. Additional criteria, including network proximity score, enrichment of GSEA gene signatures from HCoV transcriptome datasets, drug CMAP signature, and literature evidence of drug antiviral efficacy and side-effect profile, resulted in 16 candidate drugs, including anti-inflammatory, immunosuppressant and antineoplastic agents, estrogen receptor modulators and angiotensin receptor blockers. This experimentally validated drug-target network approach was readily adapted by the group in response to the COVID-19 outbreak [60–62], has multiple levels of validation for candidate drug identification and incorporates transcriptome validation as well as a protein-network focus. Drug repurposing efforts that integrate multi-omic signatures with structure data [64] and machine learning approaches have successfully identified drug candidates with anti-SARS-CoV-2 activity in vitro [65].

Overall, signature analysis using CMap, CLUE and LINCS datasets and user-developed pipelines have been adapted by many different groups to identify hundreds of potential candidate repurposable drugs [66–68]. By prioritizing gene input for disease signature generation based on pathway analyses and gene-set enrichment analyses from infected tissues or allowing virus-host and drug-target associations (DrugBank, ChEMBL, IUPHAR) to inform the selection of drugs, signature-based approaches can be adapted for targeted (e.g., specific drug class or biological pathway), biologically relevant candidate repurposable drug identification. Applying different filtering criteria (signature similarity scores, FDA approval, drug mechanism of action, manual curation of drugs in clinical trial) can also improve stringency of candidate drug identification.

It should be noted that while a number of candidate repurposable drugs show promise as antivirals in vitro [43, 69, 70] and many more are under investigation in clinical trial, safety concerns [71–73] and false positive findings [74] are common upon deeper exploration. Results from the RECOVERY (<https://www.recoverytrial.net/results>) and SOLIDARITY [75] platform clinical trials of drugs that showed early promise, like hydroxychloroquine and lopinavir, have since proven to have little or no effect in reducing mortality rates or treating COVID-19 [76].

APPLICATION TO DRUG REPURPOSING FOR CNS DISORDERS

Even in the case of COVID-19, where we have knowledge of the causative pathogen and sufficient understanding of the molecular mechanisms of infection to identify the viral proteins that can be targeted to inhibit viral replication, development of approved treatments is a challenge. For CNS diseases, a major stumbling block to the identification of effective therapies remains our limited understanding of the etiology of these disorders. The heterogeneity and complexity of CNS disorders undoubtedly contributes to the very low rates of clinical approval of new drugs, which often fail in clinical trials due to a lack of efficacy [77]. For drugs that attempt to treat the causes of neurodegenerative diseases like Alzheimer's disease and Parkinson's disease, the clinical failure rate has been high [78]. The shortcomings of animal models [79] and difficulties in designing drugs that have the physicochemical properties to be successfully delivered across the blood brain barrier are other limiting factors [80]. The challenges to successful drug discovery and drug repurposing in the CNS have been reviewed in detail elsewhere [78, 81–84].

A standard approach to discover new drugs, high-throughput screening to identify compounds that show activity against a single target or pathway, has resulted in diminishing success over the years [85], and has been particularly limited for brain-related

disorders. For example, for Alzheimer's disease the "amyloid hypothesis" led to the identification of a target pathological pathway and a search for drugs that can inhibit beta-amyloid aggregation. This has not resulted in successful disease-modifying treatment [86]. Drug repositioning using signature-based and machine-learning approaches offers an alternative strategy for drug identification for brain-related disorders, which to date has primarily relied on serendipitous findings [87] and incremental improvements of existing drugs for treatment [78].

Signature-based strategies are already identifying promising drug leads for different brain disorders. Utilizing a combination of over 60,000 CMap and L1000 drug signatures, and transcriptomic and proteomic Alzheimer's disease signatures, Lee et al devised a "drug repositioning perturbation score" to measure the inverse association between drug and disease signatures, weighted by pharmacogenomic knowledge, to identify candidate repurposable drugs with high score in both transcriptomic and proteomic data [88]. Top findings included inhibitors of monoamine oxidase, which are implicated in the pathology of Alzheimer's disease [89]. This approach has the advantage of incorporating different "omics" data types to inform drug discovery. The CMap was also deployed to identify candidate repurposable drugs for bipolar disorder, using an approach that focused on similarity with current indicated drugs for bipolar disorder rather than discordance with disease signature. Common differentially expressed genes, identified by RNAseq analysis of models, a neuronal cell line (NT2-N), and rats treated with a combination of bipolar disorder drugs (lithium, valproic acid, quetiapine, lamotrigine), were used to generate the input drug gene signatures. CMap drug signatures of 10 drugs that were concordant with the signatures of known bipolar disorder drug treatments were identified [90]. As an alternative to "omics" to generate disease signatures, Sullivan et al experimentally identified enzymes associated with bioenergetic deficits in postmortem schizophrenia brain, followed by *in silico* confirmation of these target or "seed genes" in other postmortem datasets [91]. In iLINC, gene knockdown signatures for each seed gene were generated. Next, drugs with gene signatures that were highly discordant to the seed gene signatures were identified, converging on PPAR agonists as candidate drugs of interest. Administering the PPAR-gamma agonist pioglitazone to GluN1 knockdown mice reversed cognitive deficits associated with this phenotype and was proposed as a target drug of interest for treatment of cognitive symptoms associated with schizophrenia and other neuropsychiatric conditions [91]. Taking a different approach to generating disease signatures, a multi-psychiatric disorder analysis generated signatures from genome wide association study (GWAS)-imputed transcriptome profiles [92]. GWAS-imputed transcriptome data captures the genetically regulated part of expression and overcomes some limitations of transcriptome-only derived signatures, including circumventing confounding factors like medication effects that are associated with signatures generated from postmortem tissue datasets. Applying drug-set enrichment analysis generally supported the validity of the resulting drug repurposing findings. For example, known antipsychotic drugs were enriched in schizophrenia and bipolar disorder analyses and known antidepressants were among the top findings for MDD. GWAS was also used to generate gene-set associations with disease phenotype (MDD and schizophrenia) prior to integration with drug-target relationships and chemical perturbation signatures from CMAP [93, 94]. By constructing bipartite drug-target interaction networks from these analyses, this approach assesses whether identified drugs interact with multiple targets, and whether different targets show affinity for dissimilar drugs, allowing for prioritization of drug-gene sets for further exploration [95].

One of the major limiting factors when trying to understand the pathophysiology of psychiatric disorders and identify new therapeutics is disease heterogeneity. One approach taken to

address this issue is exploring postmortem transcriptome signatures generated from subjects in depressive and remission stages of MDD and non-psychiatrically ill controls [96]. Gene co-expression analysis and Bayesian network modeling identified gene modules associated with disease state (vs trait) which was then searched against CMap signatures to identify candidate repurposable drugs. A subset of the drugs identified are already in use to treat psychiatric disorders. Such an approach can be used to identify treatments that can be targeted to clinical phases of MDD.

An advantage of signature-based approaches is the flexibility to incorporate emerging "omics" analyses, which will improve our understanding of disease mechanisms. Single-cell RNAseq studies offer significant insight into complex tissues like human brain, identifying altered patterns of cell-type specific gene expression in cell populations that are associated with neurological [97–99] and psychiatric [100] disorders. Similarly, transcriptome analysis of neuronal and glial iPSCs and cerebral organoids [101–104] derived from patients with neurological and psychiatric disorders is a useful tool, with some translational potential for interrogating disease mechanisms [105] and can be incorporated into signature-based drug repurposing analyses [106]. Although still limited relative to the available transcriptomic signatures, drug signatures for the P100, the LINCS proteomic assay platform consisting of approximately 100 representative phosphopeptide probes analyzed by mass spectrometry, offers another strand of data for drug signature identification.

A number of logistical hurdles should also be considered prior to conducting *in silico* drug repurposing. Many of the analyses reviewed here require standard computing resources found in most laboratories. However, machine learning dependent approaches may require access to a computer cluster, to provide sufficient processing power. Additionally, the databases referenced here, including LINCS, PubChem, GEO, IUPHAR etc. are publicly accessible at no cost to the user, and provide convenient interfaces for scientists to conduct analyses. However, some databases, like DrugBank, may require the creation of an account and verification of academic status prior to accessing their complete database.

Finally, limiting candidate repurposable drugs from *in silico* screens to those that have obtained FDA approval and/or are under investigation addresses a major concern in drug development, namely, drug safety. However, further assessment of any candidate drug is essential to determine if the drug is efficacious for its purported novel use, within its established safe dosage range. In addition, a vast array of *in silico* toxicology resources are available to assess drug toxicity during the drug development process. In a recent review, Pawar et al. identified almost 1,000 data resources including chemical information, and pharmacovigilance and toxicity data that could inform on toxicology during drug discovery [107]. Projects like the "Enhancing TRANslational SAFETY Assessment through Integrative Knowledge Management (eTRANSAFE)" (<https://etransafe.eu/>) are developing the infrastructure to integrate analyses of different data types to improve translational safety assessment during drug development. Integrating computational toxicology and drug repurposing approaches will undoubtedly improve the transition of candidate repurposable drugs to the clinic, by identifying potential safety concerns earlier in the process.

In summary, a large number of bioinformatic workflows and pipelines have been developed in the last two years to address the urgent need to identify candidate repurposable drugs for COVID-19. Many of these workflows may now be adapted to drug repurposing efforts for neurological and neuropsychiatric disorders, where drug discovery efforts have stagnated. Signature-based approaches using available CMap and LINCS tools are particularly amenable to use by those without formal bioinformatic training. Although significant challenges remain to successfully translate candidate drugs from bench to bedside, advances in bioinformatic drug screening may open new avenues for identifying therapies for CNS disorders.

REFERENCES

- Martinez MA. Lack of effectiveness of repurposed drugs for COVID-19 treatment. *Front Immunol*. 2021;12:635371.
- Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov*. 2004;3:673–83.
- Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019;18:41–58.
- Vatansver S, Schlessinger A, Wacker D, Kaniskan HU, Jin J, Zhou MM, et al. Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. *Medicinal Res Rev*. 2021;41:1427–73.
- Guo Y, Zhang Y, Lyu T, Prosperi M, Wang F, Xu H, et al. The application of artificial intelligence and data integration in COVID-19 studies: a scoping review. *J Am Med Inform Assoc*. 2021;28:2050–67.
- Naseem M, Akhund R, Arshad H, Ibrahim MT. Exploring the potential of artificial intelligence and machine learning to combat COVID-19 and existing opportunities for LMIC: a scoping review. *J Prim Care Community Health*. 2020;11:2150132720963634.
- Rasheed J, Jamil A, Hameed AA, Aftab U, Aftab J, Shah SA, et al. A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. *Chaos Solitons Fractals*. 2020;141:110337.
- Rasheed J, Jamil A, Hameed AA, Al-Turjman F, Rasheed A. COVID-19 in the age of artificial intelligence: a comprehensive review. *Interdiscip Sci*. 2021;13:153–75.
- Russell SJ, Norvig P, Davis E. *Artificial Intelligence: A Modern Approach*. 3rd edition edn. Prentice Hall: Upper Saddle River, New Jersey, 2010.
- Sun J, Jeliakzova N, Chupakun V, Golib-Dzib JF, Engkvist O, Carlsson L, et al. ExCAPE-DB: an integrated large scale dataset facilitating Big Data analysis in chemogenomics. *J Cheminform*. 2017;9:17.
- Wang Z, Li J, Dang R, Liang L, Lin J. PhIN: a protein pharmacology interaction network database. *CPT Pharmacomet Syst Pharm*. 2015;4:e00025.
- Chaudhari R, Fong LW, Tan Z, Huang B, Zhang S. An up-to-date overview of computational polypharmacology in modern drug discovery. *Expert Opin Drug Discov*. 2020;15:1025–44.
- von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissner R. PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic acids Res*. 2011;39:D1060–1066. Database issue
- Masoudi-Sobhanzadeh Y, Omid Y, Amanlou M, Masoudi-Nejad A. DrugR+: a comprehensive relational database for drug repurposing, combination therapy, and replacement therapy. *Comput Biol Med*. 2019;109:254–62.
- Pilarczyk M, Kouril M, Shamsaei B, Kouril M, Vasiliauskas J, Niu W, Shamsaei B, et al. Connecting omics signatures of diseases, drugs, and mechanisms of actions with iLINC. *bioRxiv*; 2019. <https://doi.org/10.1101/826271>.
- Koleti A, Terryn R, Stathias V, Chung C, Cooper DJ, Turner JP, et al. Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res*. 2018;46:D558–D566.
- O'Donovan SM, Imami A, Eby H, Henkel ND, Creeden JF, Asah S, et al. Identification of candidate repurposable drugs to combat COVID-19 using a signature-based approach. *Sci Rep*. 2021;11:4495.
- Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res*. 2014;42:D1091–1097. Database issue
- Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*. 2008;36:D901–906. Database issue
- Pinzi L, Tinivella A, Caporuscio F, Rastelli G. Drug repurposing and polypharmacology to fight SARS-CoV-2 through inhibition of the main protease. *Front Pharmacol*. 2021;12:636989.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein Data Bank. *Nucleic Acids Res*. 2000;28:235–42.
- Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res*. 2021;49:D437–D451. D1
- Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020;583:459–68.
- Abd-Alrazaq A, Alajlani M, Alhuwail D, Schneider J, Al-Kuwari S, Shah Z, et al. Artificial intelligence in the fight against COVID-19: scoping review. *J Med Internet Res*. 2020;22:e20756.
- Kaushal K, Sarma P, Rana SV, Medhi B, Naithani M. Emerging role of artificial intelligence in therapeutics for COVID-19: a systematic review. *J Biomol Struct Dynam*. 2020;1–16. Online ahead of print.
- Hunter P, Chapman T, Coveney PV, de Bono B, Diaz V, Fenner J, et al. A vision and strategy for the virtual physiological human: 2012 update. *Interface Focus*. 2013;3:20130004.
- Fenner JW, Brook B, Clapworthy G, Coveney PV, Feipel V, Gregersen H, et al. The EuroPhysiome, STEP and a roadmap for the virtual physiological human. *Philos Trans A Math Phys Eng Sci*. 2008;366:2979–99.
- Clermont G, Bartels J, Kumar R, Constantine G, Vodovotz Y, Chow C. In silico design of clinical trials: a method coming of age. *Crit Care Med*. 2004;32:2061–70.
- Wang S, Sun Q, Xu Y, Pei J, Lai L. A transferable deep learning approach to fast screen potential antiviral drugs against SARS-CoV-2. *Brief Bioinform*. 2021;22:1–11.
- Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol Inf*. 2020;39:e2000028.
- Karki N, Verma N, Trozzi F, Tao P, Kraka E, Zoltowski B. Predicting potential SARS-CoV-2 drugs-in depth drug database screening using deep neural network framework SSnet, classical virtual screening and docking. *Int J Mol Sci*. 2021;22:1573–90.
- Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet*. 2020;395:e30–e31.
- Stebbing J, Krishnan V, de Bono S, Ottaviani S, Casalini G, Richardson PJ, et al. Mechanism of baricitinib supports artificial intelligence-predicted testing in COVID-19 patients. *EMBO Mol Med*. 2020;12:e12697.
- Kalil AC, Patterson TF, Mehta AK, Tomashek KM, Wolfe CR, Ghazaryan V, et al. Baricitinib plus remdesivir for hospitalized adults with Covid-19. *N. Engl J Med*. 2021;384:795–807.
- Zeng X, Song X, Ma T, Pan X, Zhou Y, Hou Y, et al. Repurpose open data to discover therapeutics for COVID-19 using deep learning. *J Proteome Res*. 2020;19:4624–36.
- Group RC, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in hospitalized patients with Covid-19. *N. Engl J Med*. 2021;384:693–704.
- Tan F, Yang R, Xu X, Chen X, Wang Y, Ma H, et al. Drug repositioning by applying 'expression profiles' generated by integrating chemical structure similarity and gene semantic similarity. *Mol Biosyst*. 2014;10:1126–38.
- Keenan AB, Jenkins SL, Jagodnik KM, Koplev S, He E, Torre D, et al. The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst*. 2018;6:13–24.
- Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*. 2017;171:1437–1452 e1417.
- Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006;313:1929–35.
- Duan Q, Reid SP, Clark NR, Wang Z, Fernandez NF, Rouillard AD, et al. L1000CDS (2): LINCS L1000 characteristic direction signatures search engine. *NPJ. Syst Biol Appl*. 2016;2:1–12.
- Islam T, Rahman MR, Aydin B, Beklen H, Arga KY, Shahjaman M. Integrative transcriptomics analysis of lung epithelial cells and identification of repurposable drug candidates for COVID-19. *Eur J Pharmacol*. 2020;887:173594.
- Le BL, Andreoletti G, Oskotsky T, Vallejo-Gracia A, Rosales R, Yu K, et al. Transcriptomics-based drug repositioning pipeline identifies therapeutic candidates for COVID-19. *Sci Rep*. 2021;11:12310.
- Jia Z, Song X, Shi J, Wang W, He K. Transcriptome-based drug repositioning for coronavirus disease 2019 (COVID-19). *Pathog Dis*. 2019;78:1–7.
- Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Moller R, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*. 2020;181:1036–1045 e1039.
- Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med*. 2020;26:842–4.
- Li X, Yu J, Zhang Z, Ren J, Peluffo AE, Zhang W, et al. Network bioinformatics analysis provides insight into drug repurposing for COVID-19. *Med Drug Discov*. 2021;10:100090.
- Bekerman E, Neveu G, Shulla A, Brannan J, Pu SY, Wang S, et al. Anticancer kinase inhibitors impair intracellular viral trafficking and exert broad-spectrum antiviral effects. *J Clin Investig*. 2017;127:1338–52.
- Schor S, Einav S. Repurposing of kinase inhibitors as broad-spectrum antiviral drugs. *DNA Cell Biol*. 2018;37:63–69.
- Mousavi SZ, Rahmanian M, Sami A. A connectivity map-based drug repurposing study and integrative analysis of transcriptomic profiling of SARS-CoV-2 infection. *Infect Genet Evol*. 2020;86:104610.
- Li Z, Yang L. Underlying mechanisms and candidate drugs for COVID-19 based on the connectivity map database. *Front Genet*. 2020;11:558557.
- Basha O, Shpringer R, Argov CM, Yeager-Lotem E. The DifferentialNet database of differential protein-protein interactions in human tissues. *Nucleic Acids Res*. 2018;46:D522–D526. D1

53. Imami AS, O'Donovan SM, Creeden JF, Wu X, Eby H, McCullumsmith CB, et al. Oxytocin's anti-inflammatory and proimmune functions in COVID-19: a transcriptomic signature-based approach. *Physiol Genomics*. 2020;52:401–7.
54. Creeden JF, Imami AS, Eby HM, Gillman C, Becker, KN, Reigle J, et al. Fluoxetine as an anti-inflammatory therapy in SARS-CoV-2 infection. *Biomedicine & Pharmacotherapy*. (2021);138:1–6. <https://doi.org/10.1016/j.biopha.2021.111437>.
55. El-Hachem N, Eid E, Nemer G, Dbaibo G, Abbas O, Rubeiz N, et al. Integrative transcriptome analyses empower the anti-COVID-19 drug arsenal. *iScience*. 2020;23:101697.
56. Southan C, Sharman JL, Benson HE, Faccenda E, Pawson AJ, Alexander SP, et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. *Nucleic Acids Res*. 2016;44:D1054–1068. D1
57. Jia Z, Liu Y, Guan N, Bo X, Luo Z, Barnes MR. Cogena, a novel tool for co-expressed gene-set enrichment analysis, applied to drug repositioning and drug mode of action discovery. *BMC Genomics*. 2016;17:414.
58. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Disco*. 2020;6:14.
59. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 2012;40:D1100–1107. Database issue
60. Cheng F, Lu W, Liu C, Fang J, Hou Y, Handy DE, et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat Commun*. 2019;10:3476.
61. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabasi AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun*. 2018;9:2691.
62. Cheng F, Kovacs IA, Barabasi AL. Network-based prediction of drug combinations. *Nat Commun*. 2019;10:1197.
63. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol*. 2012;8:e1002503.
64. Belyaeva A, Cammarata L, Radhakrishnan A, Squires C, Yang KD, Shivashankar GV, et al. Causal network models of SARS-CoV-2 expression and aging to identify candidates for drug repurposing. *Nat Commun*. 2021;12:1024.
65. Han N, Hwang W, Tzelepis K, Schmerer P, Yankova E, MacMahon M, et al. Identification of SARS-CoV-2-induced pathways reveals drug repurposing strategies. *Sci Adv*. 2021;7:1–14.
66. Loganathan T, Ramachandran S, Shankaran P, Nagarajan D, Mohan SS. Host transcriptome-guided drug repurposing for COVID-19 treatment: a meta-analysis based approach. *PeerJ*. 2020;8:e9357.
67. Hoagland DA, Clarke DJB, Moller R, Han Y, Yang L, Wojciechowicz ML, et al. Modulating the transcriptional landscape of SARS-CoV-2 as an effective method for developing antiviral compounds bioRxiv 2020: <https://doi.org/10.1101/2020.07.12.199687>.
68. Daamen AR, Bachali P, Owen KA, Kingsmore KM, Hubbard EL, Labonte AC, et al. Comprehensive transcriptomic analysis of COVID-19 blood, lung, and airway. *Sci Rep*. 2021;11:7052.
69. Huang CT, Chao TL, Kao HC, Pang YH, Lee WH, Hsieh CH, et al. Enhancement of the IFN-beta-induced host signature informs repurposed drugs for COVID-19. *Heliyon*. 2020;6:e05646.
70. Raymonda MH, Ciesla JH, Monaghan M, Leach J, Asantewaa G, Smorodintsev-Schiller LA, et al. Pharmacologic profiling reveals lapatinib as a novel antiviral against SARS-CoV-2 in vitro. Preprint at. 2020; <https://www.biorxiv.org/content/10.1101/2020.11.25.398859v1.full.pdf>.
71. Khalil A, Kamar A, Nemer G. Thalidomide-revisited: are COVID-19 patients going to be the latest victims of yet another theoretical drug-repurposing? *Front Immunol*. 2020;11:1248.
72. Ellinger B, Bojkova D, Zaliani A, Cinatl J, Claussen C, Westhaus S, et al. A SARS-CoV-2 cytopathicity dataset generated by high-content screening of a large drug repurposing collection. *Sci Data*. 2021;8:70.
73. Xing J, Shankar R, Drelich A, Paithankar S, Chekalin E, Dexheimer T et al. Analysis of Infected Host Gene Expression Reveals Repurposed Drug Candidates and Time-Dependent Host Response Dynamics for COVID-19. bioRxiv. 2020; <https://doi.org/10.1101/2020.04.07.030734>.
74. Jang WD, Jeon S, Kim S, Lee SY. Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay. *Proc Natl Acad Sci USA*. 2021;118:1–9.
75. Consortium WHOST, Pan H, Peto R, Henao-Restrepo AM, Preziosi MP, Sathiyamoorthy V, et al. Repurposed antiviral drugs for Covid-19 - interim WHO solidarity trial results. *N Engl J Med*. 2021;384:497–511.
76. Venkatesan P. Repurposing drugs for treatment of COVID-19. *Lancet Respir Med*. 2021;9:e63.
77. McArthur RA. Aligning physiology with psychology: translational neuroscience in neuropsychiatric drug discovery. *Neurosci Biobehav Rev*. 2017;76:4–21. Pt A
78. Gribkoff VK, Kaczmarek LK. The need for new approaches in CNS drug discovery: why drugs have failed, and what can be done to improve outcomes. *Neuropharmacology*. 2017;120:11–19.
79. Planchez B, Surget A, Belzung C. Animal models of major depression: drawbacks and challenges. *J Neural Transm*. 2019;126:1383–408.
80. Pajouhesh H, Lenz GR. Medicinal chemical properties of successful central nervous system drugs. *NeuroRx: J Am Soc Exp NeuroTherapeutics*. 2005;2:541–53.
81. Shukla R, Henkel ND, Alganem K, Hamoud AR, Reigle J, Alnafisah RS, et al. Signature-based approaches for informed drug repurposing: targeting CNS disorders. *Neuropsychopharmacology*. 2021;46:116–30.
82. Lee HM, Kim Y. Drug repurposing is a new opportunity for developing drugs against neuropsychiatric disorders. *Schizophr Res Treat*. 2016;2016:6378137.
83. Morofujii Y, Nakagawa S. Drug development for central nervous system diseases using in vitro blood-brain barrier models and drug repositioning. *Curr Pharm Des*. 2020;26:1466–85.
84. Kaiser T, Feng G. Modeling psychiatric disorders for developing effective treatments. *Nat Med*. 2015;21:979–88.
85. Booth B, Zimmel R. Prospects for productivity. *Nat Rev Drug Discov*. 2004;3:451–6.
86. Mangialasche F, Solomon A, Winblad B, Mecocci P, Kivipelto M. Alzheimer's disease: clinical trials and drug development. *Lancet Neurol*. 2010;9:702–16.
87. Lopez-Munoz F, Alamo C. Monoaminergic neurotransmission: the history of the discovery of antidepressants from 1950s until today. *Curr Pharm Des*. 2009;15:1563–86.
88. Lee SY, Song MY, Kim D, Park C, Park DK, Kim DG, et al. A proteotranscriptomic-based computational drug-repositioning method for Alzheimer's disease. *Front Pharmacol*. 2019;10:1653.
89. Quartey MO, Nyarko JNK, Pennington PR, Heistad RM, Klassen PC, Baker GB, et al. Alzheimer disease and selected risk factors disrupt a co-regulation of monoamine oxidase-A/B in the hippocampus, but not in the cortex. *Front Neurosci*. 2018;12:419.
90. Kidnapillai S, Bortolasci CC, Udawela M, Panizzutti B, Spolding B, Connor T, et al. The use of a gene expression signature and connectivity map to repurpose drugs for bipolar disorder. *World J Biol Psychiatry*. 2020;21:775–83.
91. Sullivan CR, Mielnik CA, O'Donovan SM, Funk AJ, Bentea E, DePasquale EA, et al. Connectivity analyses of bioenergetic changes in Schizophrenia: identification of novel treatments. *Mol Neurobiol*. 2019;56:4492–517.
92. So HC, Chau CK, Chiu WT, Ho KS, Lo CP, Yim SH, et al. Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat Neurosci*. 2017;20:1342–9.
93. Gaspar HA, Breen G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci Rep*. 2017;7:12460.
94. Gaspar HA, Gerring Z, Hubel C, Major Depressive Disorder Working Group of the Psychiatric Genomics C, Middeldorp CM, Derks EM, et al. Using genetic drug-target networks to develop new drug hypotheses for major depressive disorder. *Transl Psychiatry*. 2019;9:117.
95. Maggiora G, Gokhale V. A simple mathematical approach to the analysis of polypharmacology and polyspecificity data. *F1000Res* 2017;6.
96. Shukla R, Newton DF, Sumitomo A, Zare H, McCullumsmith R, Lewis DA et al. Molecular characterization of depression trait and state. *Mol Psychiatry* 2021. <https://doi.org/10.1038/s41380-021-01347-z>.
97. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570:332–7.
98. Olah M, Menon V, Habib N, Taga MF, Ma Y, Yung CJ, et al. Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nat Commun*. 2020;11:6129.
99. Leng K, Li E, Eser R, Piergies A, Sit R, Tan M, et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat Neurosci*. 2021;24:276–87.
100. Nagy C, Maitra M, Tanti A, Suderman M, Theroux JF, Davoli MA, et al. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nat Neurosci*. 2020;23:771–81.
101. Lin M, Pedrosa E, Hrabovsky A, Chen J, Puliafito BR, Gilbert SR, et al. Integrative transcriptome network analysis of iPSC-derived neurons from schizophrenia and schizoaffective disorder patients with 22q11.2 deletion. *BMC Syst Biol*. 2016;10:105.
102. Sawada T, Chater TE, Sasagawa Y, Yoshimura M, Fujimori-Tonou N, Tanaka K, et al. Developmental excitation-inhibition imbalance underlying psychoses revealed by single-cell analyses of discordant twins-derived cerebral organoids. *Mol Psychiatry*. 2020;25:2695–711.
103. Kathuria A, Lopez-Lengowski K, Jagtap SS, McPhie D, Perlis RH, Cohen BM, et al. Transcriptomic landscape and functional characterization of induced pluripotent stem cell-derived cerebral organoids in schizophrenia. *JAMA Psychiatry*. 2020;77:745–54.

104. Kathuria A, Lopez-Lengowski K, Vater M, McPhie D, Cohen BM, Karmacharya R. Transcriptome analysis and functional characterization of cerebral organoids in bipolar disorder. *Genome Med.* 2020;12:34.
105. Hoffman GE, Hartley BJ, Flaherty E, Ladran I, Gochman P, Ruderfer DM, et al. Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. *Nat Commun.* 2017;8:2225.
106. Readhead B, Hartley BJ, Eastwood BJ, Collier DA, Evans D, Farias R, et al. Expression-based drug screening of neural progenitor cells from individuals with schizophrenia. *Nat Commun.* 2018;9:4412.
107. Pawar G, Madden JC, Ebbrell D, Firman JW, Cronin MTD. In silico toxicology data resources to support read-across and (Q)SAR. *Front Pharmacol.* 2019;10:561.

ACKNOWLEDGEMENTS

Funding sources: RO1 MH107487, RO1 MH121102, RO1 AG057598.

AUTHOR CONTRIBUTIONS

ASI, REM and SMOD were involved in the conceptualization and writing of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Sinead M. O'Donovan.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021