Title
Insights into Computational Drug Repurposing for Neurodegenerative Disease

Journal
Trends in Pharmacological Sciences, 40(8)

ISSN
0165-6147

Authors
Paranjpe, Manish D
Taubes, Alice
Sirota, Marina

Publication Date
2019-08-01

DOI
10.1016/j.tips.2019.06.003

Peer reviewed

Published in final edited form as:

Trends Pharmacol Sci. 2019 August ; 40(8): 565–576. doi:10.1016/j.tips.2019.06.003.

Insights into Computational Drug Repurposing for Neurodegenerative Disease

Manish Paranjpe[1,*], **Ali Taubes**[2], **Marina Sirota**[1,2,*]

[1]Bakar Computational Health Sciences Institute, University of California, San Francisco, California, 94158 USA

[2]Gladstone Institutes, San Francisco CA, 94158 USA

Abstract

Computational drug repurposing has the ability to remarkably reduce drug development time and cost in an era where these factors are prohibitively high. Several examples of successful repurposed drugs exist in fields such as oncology, diabetes, leprosy, inflammatory bowel disease, among others, however computational drug repurposing in neurodegenerative disease has presented several unique challenges stemming from the lack of validation methods and difficulty in studying heterogenous diseases of aging. Here, we examine existing approaches to computational drug repurposing, including molecular, clinical, and biophysical methods, and propose data sources and methods to advance computational drug repurposing in neurodegenerative disease using Alzheimer's as an example.

Keywords

artificial intelligence; machine learning; transcriptomic analysis; EHR; EMR; Alzheimer's Disease

Computational drug repurposing can be a fast and cheap tool for drug development

Neurodegenerative disease has posed significant and unique challenges to effective drug discovery over the past century. More than five million Americans are living with Alzheimer's Disease (AD), and more than 500,000 have been diagnosed with Parkinson's Disease (PD), the two most common neurodegenerative disorders[1], [2]. Millions more are suffering from rarer conditions such as Frontotemporal Dementia (FTD), Huntington's disease, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), and Spinal Muscular Atrophy (SMA), among others [3]–[6]. AD and other dementias account for over

*Correspondence: marina.sirota@ucsf.edu (Marina Sirota, PhD); manish.paranjpe@ucsf.edu (Manish Paranjpe BA).

Conflicts of Interest

Marina Sirota is a former employee of Pfizer and is a scientific advisor to twoXAR.

$200 billion in healthcare costs, a number that is projected to rise by 2050 if these diseases remain untreatable. Since 2000, nine immunomodulatory compounds have reached Food and Drug Administration (FDA) approval for MS, some of which work to slow disease progression[5], however there is no cure. This sits in stark contrast to only four non-disease modifying compounds approved for AD during that time period, even while the population of AD patients is projected to almost double in the next ten years[7]–[10]. The unmet need for the timely development of effective therapies for AD and other neurodegenerative has been steadily climbing, with the burden on our healthcare system reaching a critical level[2].

Computational drug repurposing, or the *in silico* screening of FDA-approved compounds for use against new indications, promises to get new and effective neurodegenerative disease therapies to the clinic faster [11], [12]. Traditional drug development entails discovery and pre-clinical research, safety review, clinical studies, FDA review, and FDA post-market safety monitoring, and can take 15 years and over one billion dollars to bring a drug to market[13],[14]. Many repurposed drugs on the other hand, have already been FDA approved and therefore face a cheaper and quicker journey to the clinic[13], [15]. Developments in high throughput screening technologies, along with the growing repository of ‘omics-based data across disease indications, has catapulted computational drug repurposing methods to the forefront of attractive drug discovery techniques for neurodegenerative disease[13], [16], [17]. While computational drug repurposing has the potential to greatly reduce drug development time and cost, experimental and economic obstacles must be overcome.

Existing computational approaches to drug repurposing

Previously, successful repurposed drugs came from serendipitous events in the lab and clinic. One relevant example of this is the use of zonisamide for the treatment of PD[18], [19]. Upon using zonisamide to treat a Japanese epilepsy patient with PD, Murata discovered that zonisamide improved the patient’s PD symptoms too. Based on this serendipitous finding, zonisamide was approved as an anti-PD therapy in 2009 in Japan. Recently, the emergence of high throughput molecular, clinical and structural biology technologies combined with the advent of economically feasible large-scale computational capacity has created a novel opportunity to rationally repurpose existing drugs using computational frameworks instead of chance findings. Existing computational approaches to drug repurposing can be divided into molecular, clinical, and structure-based (biophysical) methods (Table 1). Here, we highlight the advantages and current limitations of each strategy, highlight key innovations and propose future solutions in the context of neurodegenerative disease.

Molecular methods of drug repurposing

Our group has pioneered the development of transcriptome-based approaches to computational drug discovery[20] which aim to compare drug gene expression signatures pre-and post-drug treatment to disease gene expression signatures in order to predict drugs that may reverse disease gene signatures[21]. The early iteration of this hypothesis takes the form of the Connectivity Map (CMap) [i], a resource created at the Broad Institute which

consists of a database of microarray gene expression signatures that have been constructed pre and postapplication of >1200 small molecule compounds to cell lines [22] [22]. Recent examples of success from this method include preclinical identification of drugs such as parbendazole for osteoporosis [23], cimetidine for lung adenocarcinoma [20], citalopram for metastatic colon cancer[24], and topiramate for inflammatory bowel disease [25] (Table 1). More recently developed methods combine genetic and gene expression data and employ network-based approaches to advance drug discovery[26]. Such methods enable high-throughput screening of existing compounds, and do not require *a priori* identification of target molecule. However, transcriptomic methods are limited by the availability of subtype-specific disease expression signatures (Table 1). This is particularly relevant in the case of heterogeneous neurodegenerative diseases such as AD and PD[27]. Recent work has shown differences in AD disease pathology, biomarker profiles and disease risk due to presence of several genetic, demographic and lifestyle-related risk factors, including sex[28]–[31], the apolipoprotein E (ApoE) ε4 allele [32]–[3 7], and vascular risk factors such as diabetes and hypertension[38]–[40]. The 2018 National Institute of Aging (NIH) and Alzheimer's Association framework, which moves from classifying AD as a clinical syndrome with or without evidence of neuropathologic change towards a biomarker-based classification, is a promising step forward towards developing biologically-defined disease subtypes for AD[41]. By framing AD as a biological disease marked by distinct biomarker changes, transcriptomic methods can focus on creating drug predictions in biologically-defined subtypes. While the NIH-Accelerating Medicines Partnership – AD (NIH AMP-AD) portal [ii] has paved the way for aggregating large scale neurodegenerative disease-associated datasets for easy research access, more resources, including single cell RNA-seq datasets for the targeted development of drug predictions in disease-relevant brain cell types and, larger datasets with available genetic information are sorely needed.

Commonly used transcriptomic methods leveraging public databases such as CMap[22] or the scaled up NIH Library of Integrated Network-Based Cellular Signatures (LINCS) program[42][iii] primarily use cancer cell lines as *in-vitro* models to construct drug expression signatures by systematically applying a drug or vehicle to the cell line and using high throughput transcriptomic assays such as RNA-sequencing to profile gene expression. However, several studies have demonstrated significant gene expression differences between culture models and freshly prepared biopsy data, especially in central nervous system (CNS) tissues[43]–[46]. Therefore, while resources such as CMap and LINCS may be well suited for cancer drug discovery, they are currently limited in the case of neurodegenerative disease. Ultimately, there is a large need for the creation of transcriptomic drug perturbation databases in CNS tissues such that drug response can be measured in relevant tissue and cell types for neurodegenerative disease (Table 1; Figure 1).

Recent and promising advances in molecular-based drug repurposing techniques also include multi-omics techniques that integrate several types of molecular data such as genetic, transcriptomic, proteomic, and metabolomic data to develop computational drug

i) https://www.broadinstitute.org/connectivity-map-cmap
ii) https://www.nia.nih.gov/research/amp-ad
iii) http://www.lincsproject.org/

predictions [47], [48] One such example in the case of neurodegenerative disease is the work by Zhang et al [47] where the National Human Genome Research Institute-European Bioinformatics Institute Genome Wide Association Study (NHGRI-EBI GWAS) catalog[iv][49], PubMed, and the Human Metabolome Database [50] were systematically mined to create a collection of proteomic, metabolomic and genetic signatures of AD. By integrating this multi omics data with the Therapeutic Target Database[51] and DrugBank[52] drug-target databases, the study authors were able to generate a list of 75 drug predictions in AD. Methods incorporating several types of omics data are particularly promising in the case of CNS diseases, such as AD and PD, that have unclear genetic and epigenetic etiology.

Clinical methods of drug repurposing in neurodegeneration and beyond

Clinical methods of drug discovery leverage large-scale health data such as Electronic Medical Record (EMR), insurance claims data, clinical trial data, health registries and health surveys, personal genome testing companies. Prominent examples of EMR databases are the Mt. Sinai Bio*Me* [16] cohort and the eMERGE network [53]. Analysis of large health repositories such as EMR and claims data are particularly well suited for precision medicine as, with a sufficiently large discovery data set, researchers can identify drugs that work in distinct patient populations (Table 1). By taking advantage of patient medication histories, drugs can be identified that are effective for indications other than the primary drug use. For example, recent reanalysis of clinical trials and Medicare pharmacy claim data has suggested that statin users experience a lower incidence of AD than their non-user counterparts[54], [55] (Table 1). Similarly, researchers took a correlation-based approach using EMR laboratory testing data from Ajou University to develop “clinical-signatures” or laboratory test values before and after a patient is administered a drug. By comparing these clinical signatures, the group found terbutaline sulfate and ursodeoxycholic acid, a treatment for Kawasaki syndrome, to elicit similar changes in laboratory values. Comparing disease pairs revealed a high degree of clinical signature similarity between Kawaskai syndrome and ALS, suggesting that terbutaline sulfate may also be efficacious in ALS. Subsequent *in-vivo* validation showed that administration of terbutaline sulfate rescued axonal growth and neuromuscular junction deficits in a zebrafish model of ALS [56](Table 1).

One of the major limitations of clinical-data based drug repurposing methods lies in the publicly inaccessibility of most large-scale EMR databases, including eMerge and the Mt. Sinai Bio*Me* initiative. Moreover, clinical data from disparate sources such as insurance claims and EMR data are often unstructured, contain different terminologies and coding formats for clinical indications and medications as well as different data formats, and also suffer from issues with sparse or incomplete data (Table 1). Researchers must therefore invest time and resources into creating structured databases that integrate disparate sources of data to support computational pipelines. Semi-structured and unstructured EMR text must be converted into a computer-readable format using natural language processing techniques before knowledge discovery.

iv) https://www.ebi.ac.uk/gwas/

Indeed much progress has been made in automatic knowledge retrieval from unstructured EMR data [57]–[59]. Recently, the Observation Health Data Sciences and Informatics program has created the **OMOP** (see Glossary) Common Data Model to transform claim data and EMR record data into a standardized data format with common data representations (ie. terminologies, coding schemes, etc)[60] (Table 1). This allows subsequent statistical analyses to be conducted with minimal information loss due to variable data coding and formatting. Other linkage techniques include probabilistic matching strategies and "fuzzy" matching techniques that use multiple field values to match records even when no single field is an exact match[61], [62]. In addition to challenges related to variable data coding, EMR data is often noisy with missing variables or miscoded diagnoses, medication orders and laboratory orders [60]. These variables must be either imputed from related variables or removed from subsequent analyses[63]–[66]. Recently, the use of a **deeply-learned autoencoder** to impute missing EMR data from the ALS Pooled Resource Open-access Clinical Trial database [v] outperformed popular existing **imputation** methods such as **mean**[67], median[67], **singular value decomposition**[68], **SoftImpute**[69], and **k-nearest neighbors**[70] in both imputation accuracy and ALS disease progression predictive ability[63]. Lastly, the development of better data anonymization methods is fundamental to the release of these invaluable datasets, and better methods need to be developed in the near future.

Neurodegenerative diseases such as PD and AD often display disease pathology years or even decades before the onset of neurogenerative processes. For example, patients with PD display non-motor symptoms such as **hyposmia,** depression, olfactory dysfunction years before the onset of motor symptoms[71]–[75]. Similarly, AD patients exhibit signs of brain pathology such as increased cerebral blood flow, glucose hypometabolism, hippocampal volume loss, β-amyloid deposition and tau neurofibrillary tangles sometimes decades before cognitive decline[76]–[82]. As such, patient cohorts must be studied longitudinally over time to accurately measure disease progression and assess the efficacy of drugs in preclinical stages of disease. Observational health data would need to include patient observation spanning several years (Table 1). Longitudinal profiling and recording end of life health outcomes is made even more challenging because patients often switch health care systems or may transfer to hospice care facilities for palliative care. One example of an exemplary comprehensive database is Norwegian Prescription Database (NorPD)[vi] that exists for residents of Norway. In 2017, Brakedal et al., successfully mined the NorPD database of Norwegian diagnosis codes to identify an association of patients on glitazone with decreased incidence of PD [84] proving that these retrospective studies are possible with access to detailed longitudinal data (Table 1). While a phase II clinical trial failed to show any disease-modifying effect of glitazone in early PD, this trial included a follow-up period of only 44 weeks and did not assess the effect of glitazone on incidence of PD [85]. Ideally, glitazone should be followed up in longer clinical trials to assess long-term neuroprotective effects.

A further limitation that must be overcome in order to enhance the usability of clinical data is the addition of supporting genetic data. In the case of highly heterogenous

v) https://nctu.partners.org/proact
vi) http://www.norpd.no/

neurodegenerative disorders such as AD or PD, drug efficacy may vary by the presence of genetic risk factors[27], and subtypes (Table 1; Figure 1). For example, the effect of statins on cognitive decline in AD has been previously shown to be more effective in ApoE ε4 carriers compared to non-carrier[87]. Large scale integration of supporting genetic data will also enable researchers to engage in precision-medicine-based analysis techniques for drug development.

Lastly, as all analysis of clinical data is inherently retrospective, drug repurposing results of clinical data must be followed by prospective clinical trials. This is exemplified by the fact that a newly-minted retrospective association that was found between increased statin use and reduced incidence of ADhas failed to be replicated in randomized controlled trials [88]–[90], demonstrating the inability of retrospective analyses to prove causation. However, given the recent progress reviewed here, we are hopeful that in the near future the size and number of genetically-linked EMR databases, in which patient undergo both clinical documentation through EMRs and genetic profiling using whole exome sequencing or microarray technologies, will continue to grow. This combined with the creation of better data cleaning and de-identification methods (in which personally identifiable information such as patient name, are removed from EMRs), will facilitate the discovery of new therapies for neurodegenerative disease.

Biophysical methods of drug repurposing in neurodegenerative disease

Biophysical methods of drug repurposing include structural, ligand-based and molecular docking methods and may be particularly useful in neurodegenerative diseases with known targets such as Huntington's disease [4], [91] (Table 1). These computationally-efficient methods leverage the biochemical properties of drugs such as binding affinity and biophysical properties such as 3D conformation to achieve drug-target predictions[92]–[94]. Structural methods, such as those that use local site similarity metrics to compare protein binding sites [95] or those that identify chemiosmotic protein environments (two protein environments that can bind the same ligand) [96], [97], leverage protein conformational information to identify structurally similar drugs that might harbor similar targets[92]. If two diseases are hypothesized to share similar target proteins, then a drug or structurally similar molecule may be effective in both diseases[98]. For example, patients with AD and Huntington's disease both have increased extrasynaptic NR2B- subunit containing N-methyl-D-aspartate receptors (NMDARs) and increased phosphorylation of NMDARs [99]–[101]. Identifying drugs that inhibit extrasynaptic NMDAR activity through acting at structurally similar ligands or binding sites represents a potential drug repurposing target for both of these conditions[102].

Ligand-based methods use chemical and biological information such as binding affinity, cellular activity and absorption, distribution, metabolism and excretion (ADME) data to identify novel targets for existing drugs[92], [103]. Ligand-based screening assumes that if two molecules share a similar bioactivity profile, then they may share similar targets. In contrast to structure-based methods, ligand-based methods do not require Nuclear Magnetic Resonance (NMR) or X-ray crystal structures and instead rely on public bioactivity databases such as PubChem, ChEMBL, and DrugBank[52], [104], [105]. These databases

contain far more bioactivity profiles than crystal structures in Protein Data Bank (PDB) [105]. However, ligand-based screening suffers from the “activity cliff problem” in which two compounds with similar bioactivity profiles often exhibit remarkably different activities [106]. As the accuracy and chemical breadth of these datasets expands, the usefulness of ligand-based screening will continue to improve. Docking based methods perform molecular docking simulations to predict either novel targets for existing drugs or potential drugs for a given target[107]. One such example of docking based repurposing is the use of high throughput ligand-protein inverse docking to identify as a droperidol as a putative drug in AD due to its high binding affinity to seven AD target proteins [108] (Table 1). By performing a large virtual screen of 1553 FDA-approved drugs, the study authors could use docking simulations to determine the free energy of binding of each of the drugs to seven AD target proteins. A major limitation to docking-based methods, similar to structure-based methods, is the need for X-ray crystal structures or NMR of the drug and target.

Overall, while biophysical methods are efficient at identifying potential drug-target interactions, they all suffer from the need of an *a priori* identification of a target molecule (Table 1). Better insights into the true mechanisms of neurodegenerative diseases is required for biophysical methods of drug repurposing to achieve success. This need is highlighted by the consistent devastating failures of all clinical trials that have targeted canonical AD targets such as amyloid-beta[109]–[113]. It is clear from these failures that the literature surrounding causative pathological pathways in AD may require review.

Incorporating artificial intelligence approaches: old data with new methods

Artificial intelligence (AI) and **machine learning (ML)** methods are particularly adept at combining disparate types of data. In recent years, there has been a burgeoning interest in developing ML techniques to effectively mine transcriptomic[114], [115], structural[116]–[118] and clinical data[119]–[121]. Indeed, recently several companies have developed AI and ML-based frameworks for drug discovery. For example, IBM used AI-based text-mining strategies to create a semantic model of ALS-associated RNA-binding proteins that may represent drug targets. By applying this model to a new set of RNA-binding proteins, IBM could uncover potential ALS-associated RNA-binding[122]. An especially attractive application of ML in the context of computational drug repurposing is in the integration of molecular and biophysical data. One such recent method of integrating biophysical and molecular data used drug-induced gene expression signatures, molecular target information and structural information as features to train a multi-class support vector machine (a type of ML model) to predict the therapeutic class (ie. calcium-channel blocker, diuretic, etc) of a given drug[123]. Potential drug repurposing instances arise when the predicted therapeutic class for a drug is different than its original therapeutic class. The framework showed a classification accuracy of 78%, demonstrating its potential usefulness in the future for neurodegenerative disease as drug-induced gene expression signature and structural databases relevant to neurodegenerative disease continue to grow. While AI and ML models have shown much promise in areas such as disease prediction for PD[124], [125], MS [126], [127], AD [128]–[130],their full utility in computational drug repurposing for neurogenerative disease would be realized as the molecular, structural and clinical data resources for neurodegenerative disease increase.

Lack of validation strategies and economic incentives limits the utility computational drug repurposing for neurodegenerative disease

A major obstacle to computational drug repurposing for neurodegenerative disease is the lack of clear validation strategies. While all *in silico* strategies for computational drug repurposing require *invitro* and *in vivo* experimental validation methods, this problem is more acute in the case of drug repurposing for complex and heterogeneous neurodegenerative disease. Unlike cancer in which much progress has been made in modeling disease using highly-specific molecular-subtype-related cell lines or patient-derived mouse xenograft models, neurodegenerative diseases suffers from a lack of validated animal and cell models (Figure 1) [131]. Mouse models of neurodegenerative diseases such as AD, PD, FTD and ALS often accurately model early-stage proteinopathies but poorly recapitulate the entire pathophysiological course of disease[132], [133]. The inability of neurodegenerative disease animal models to completely phenocopy human disease may underlie the discrepancy between successful preclinical results and failed clinical trials. For example, the widely used mouse models of AD rely on introducing several exceedingly rare autosomal dominant mutations in the amyloid processing pathway which have never naturally occurred in tandem in humans, causing a precocious amyloid phenotype leading to early cognitive impairments and amyloid pathology in the mouse brain that may not accurately represent human disease[133]. Most of the failed clinical trials over the last decade, have tested anti-amyloid compounds that while effective in these amyloid-distressed mice, are ineffective if not detrimental in humans[134]. The inability and reluctance of large pharma to move away from these biologically questionable mouse models has resulted in all AD therapies being stymied in clinical trials. As genome editing techniques continue to advance and models are validated with increasing sophistication using multi omics profiling and *in vivo* imaging, models that better phenocopy human pathologies will arise. In the meantime, investigators are urged to carefully consider the limitations in using animal models for validation of computational hits.

In contrast to animal models, *in vitro* cell culture models are a high-throughput and cost-effective way to validate computational predictions. Unfortunately, while protocols for establishing neural and glial primary culture and cell lines are well established, these models lack complex neural circuitry and neuroinflammatory processes[132]. For this reason, cell culture models and induced pluripotent stem cell (iPSC)-derived culture models are limited in their ability to validate drug efficacy in neurodegenerative diseases with complex circuit deficits and inflammation such as AD. Recently, several groups have produced 3-D cell culture from iPSC-derived neurons. While 3-D culture models better replicate human neural network dynamics[135], [136] including neuron-glial interaction, challenges remain in incorporating nonneuronal cell types such as microglia to accurately model neuroinflammatory phenotypes. In spite of these limitations, iPSC and 3-D culture technology have given us a remarkable tool to study drug response for neurodegenerative disease.

A further pitfall to computational drug repurposing is limited patent exclusivity and protection of intellectual property (Figure 1). Currently, a repurposed drug can be filed

through the 505(b)(2) section of the 1984 Drug Price Competition and Patent Term Restoration Act[vii]. The 505(b)(2) pathway allows the applicant to patent an already approved drug for a new indication, dosage, route of administration, strength or formulation. The ability to use safety and efficacy data on the active ingredient from studies not conducted by the applicant makes 505(b)(2) an ideal path for repurposed drugs. However, New Drug Applications (NDA) with an existing active patent have an average of only 9.2 years of patent life left [137]. This limited window makes it difficult for applicants to generate sufficient data to patent a secondary indication for the drug. Even if a repurposed drug is approved, payers and providers may be reluctant to reimburse for the repurposed drug if an off-label or especially generic drug is already available [138]. Further, unlike NDA for new drugs that are granted 5 years of exclusivity, 505(2)(b) repurposed drugs are only given 3 years of exclusivity [110]. This further limits the return on investment for repurposed drugs. In the context of neurodegenerative disease, repurposed drugs require expensive and lengthy clinical trials to prove efficacy in cases of diseases with long disease courses such as AD. Repurposed drugs may also be required to demonstrate blood-brain barrier penetrance for use in diseases of the CNS. Moreover, elderly patient cohorts with neurodegenerative diseases may be already prescribed several additional medications that interact with the drug of interest. This will require the repurposed drug to undergo additional tolerance experiments[138], further increasing drug development time and diminishing return on investment.

Conclusions and future perspectives

Computational drug repurposing strategies have the potential to dramatically reduce the time and cost of developing therapies for neurodegenerative disease. The advent of large-scale computing capacity combined with the creation of public transcriptomic databases such as CMap, large de-identified EMR databases, and large-scale structural data represent a unique opportunity to apply computational methods to discover novel drugs for neurodegenerative indications. In spite of these resources, several economic and scientific obstacles must still be overcome (see Outstanding Questions). We see three principal problems in the field: (1), lack of subtype and tissue-specific discovery data, (2) lack of validation methods and (3) lack of pharmaceutical investment in low-return-on-investment drug repurposing ventures (Figure 1). To address the first issue, we propose the creation of database similar to the CMap and LINCS programs using CNS-derived tissue. Further we believe that it is the requirement of the hour to invest time and resources in creating more neurodegenerative disease-relevant animal models, as this has been a significant obstacle to success in finding new therapeutic compounds. Most recently, the MODEL-AD consortium, a collaborative effort between Indiana University, Jackson Laboratory, Sage Bionetworks, and University of California Irvine, has made remarkable progress in modeling sporadic, late-onset AD in rodents [139]. For example, the consortium has already created transgenic mice models expressing mutant forms of ApoE[140], Trem2[141]. Rodent models of AD have been criticized for their inability to fully recapitulate human phenotypes and drug response. In response, the IMPRiND consortium is developing a macaque model of Alzheimer’s disease

vii) https://www.govinfo.gov/content/pkg/STATUTE-98/pdf/STATUTE-98-Pg1585.pdf

while researchers at the RIKEN institute are developing a PSEN-mutation-based marmoset model of AD [142].

Thirdly, we urge the field to increase access and availability of genetically-linked EMR databases. In addition to EMR data, failed clinical trial data represents a rich dataset for computational drug discovery applications. While much progress has been in releasing patient-level data from clinical trials, most journals still do not require the sharing of patient-level de-identified data as a precondition to publication [143]. Further, consent for sharing patient-level clinical trial data is not routinely obtained from clinical trial participants. Journal policies surrounding disclosure of patient-level data and study consent protocols must be made unified and standardized (Figure 1). Lastly, funding for drug repurposing efforts, especially from non-profit organizations, is of utmost importance, and will ensure that better drugs for neurodegenerative diseases can reach the clinic faster (Figure 1). Several noteworthy nonprofits, such as the Michael J. Fox Foundation for Parkinson's Research[viii], the Alzheimer's Drug Discovery Foundation[ix], Multiple Myeloma Research Foundation[x] Cures Within Reach[xi] and the NIH have created funding opportunities to support the development of tools and animal models for computational drug repurposing [138]. As of June 2019, the National Center for Advancing Translational Science (NCATS) has three funding opportunities available (PAR-19-909[xii], PAR-18-910[xiii] and PAR-14-213[xiv]) for research in repurposing drugs. While these two pilot grants are credible first attempts, other NIH departments such as National Institute on Aging should create separate funding opportunities for neurodegenerative disease.

While AI and ML technologies have already led to outstanding advances in areas such as using imaging to predict disease progression in diseases such as AD, their utility in drug repurposing requires the creation of the above resources. AI and ML models are particularly useful for their ability to integrate diverse sources of data and identify complex patterns. However, these approaches are only as good as the datasets they are using and fall prey to the same pitfalls as more traditional transcriptomic, biophysical and clinical approaches. As better and more accurate computing methods continue to be developed, the need for better and more accessible databases across data modalities becomes ever more acute. Ultimately, following the creation of resources and initiatives specific to neurodegenerative diseases and the application of effective computational methods, computational drug repurposing has the potential to become a viable strategy for bringing to market cost-effective and timely therapies for these devastating conditions.

Acknowledgements and Funding

The authors thank members of the Sirota lab, especially Ishan Paranjpe, for their helpful discussions. MS and MP are supported by NIA 1R0AG057683 and R01AG060393.

viii) https://www.michaeljfox.org/
ix) https://www.alzdiscovery.org/
x) https://themmrf.org/
xi) https://www.cureswithinreach.org/
xii) https://grants.nih.gov/grants/guide/pa-files/PAR-18-909.html
xiii) https://grants.nih.gov/grants/guide/pa-files/PAR-18-910.html
xiv) https://grants.nih.gov/grants/guide/rfa-files/RFA-TR-17-001.html

GLOSSARY

Artificial intelligence:
A field of computer science that develops computer systems which mimic human ability to problem solve and learn. Examples of artificial intelligence relevant to computational drug repurposing include machine learning.

Deeply-learned autoencoder:
A type of artificial intelligence model that seeks to determine efficient data encodings to represent data in a lower dimension and then reconstruct the original data. Denoising autoencoders introduce noise by randomly setting some of the input values to be missing and learn to reconstruct an uncorrupted version of the data. A denoising autoencoder can be trained to reconstruct EMR data with from data with missing values

Hyposmia:
reduced ability to smell

Imputation:
replacing missing data with estimated values

k-nearest neighbors imputation:
An imputation technique that replaces missing data values with values predicted using the k-nearest neighbors algorithm

Machine learning:
A subfield of artificial intelligence that seeks to create a mathematical model based on input features (ie. gene expression, lab values, etc) and outcome states (disease vs control) in training data. The mathematical model can be used to predict outcomes or classify samples in new data.

OMOP:
A universal scheme to transform claim data and EMR record data into a standardized data format with common data representations (ie. terminologies, coding schemes, etc)

Mean imputation:
an imputation technique in which missing data for a given variable is replaced with mean of all values for that variable

Median imputation:
an imputation technique in which missing data for a given variable is replaced with median of all values for that variable

Singular value decomposition imputation:
an imputation technique that first performs mean value imputation on an input matrix. Then, singular value decomposition is performed to develop a set of mutually orthogonal vectors (termed eigengene) whose linear combination approximates the input dataset (review of singular value decomposition is beyond the scope of this article.) Data that was originally missing from the given variable is predicted by regressing that variable against the top

eigengenes and then using the coefficients of the regression to predict the missing data value.

SoftImpute:
an imputation technique that iteratively replaces the missing values in a matrix with values obtained by performing soft-thresholded SVD. The algorithm minimizes a nuclear-norm regularized loss function. A complete discussion of SoftImpute is beyond the scope of this article.

References

[1]. Karlawish J, Jack CR, Rocca WA, Snyder HM, and Carrillo MC, "Alzheimer's disease: The next frontier—Special Report 2017," Alzheimer's Dement, 2017.

[2]. Gooch CL, Pracht E, and Borenstein AR, "The burden of neurological disease in the United States: A summary report and call to action," Annals of Neurology, 2017.

[3]. Katsnelson A, De Strooper B, and Zoghbi HY, "Neurodegeneration: From cellular concepts to clinical applications," Science Translational Medicine. 2016.

[4]. McColgan P and Tabrizi SJ, "Huntington's disease: a clinical review," European Journal of Neurology. 2018.

[5]. Correale J, Gaitán MI, Ysrraelit MC, and Fiol MP, "Progressive multiple sclerosis: from pathogenic mechanisms to treatment," Brain. 2017.

[6]. Kolb SJ and Kissel JT, "Spinal muscular atrophy: A timely review," Archives of Neurology. 2011.

[7]. Crismon ML, "Tacrine: First drug approved for alzheimer's disease," Ann Pharmacother, 1994.

[8]. Birks JS and Harvey RJ, "Donepezil for dementia due to Alzheimer's disease," Cochrane Database of Systematic Reviews. 2018.

[9]. Loy C and Schneider L, "Galantamine for Alzheimer's disease and mild cognitive impairment," Cochrane Database Syst. Rev, no. 1, 1 2006.

[10]. Briks J and Grimley Evans J, "Rivastigmine for Alzheimer ' s disease.," Cochrane database Syst. Rev, 2015.

[11]. Kim TW, "Drug Repositioning Approaches for the Discovery of New Therapeutics for Alzheimer's Disease," Neurotherapeutics. 2015.

[12]. Corbett A et al., "Drug repositioning for Alzheimer's disease," Nature Reviews Drug Discovery. 2012.

[13]. Hodos RA, Kidd BA, Shameer K, Readhead BP, and Dudley JT, "In silico methods for drug repurposing and pharmacology," Wiley Interdiscip. Rev. Syst. Biol. Med, 2016.

[14]. Kaitin KI, "Deconstructing the drug development process: The new face of innovation," Clinical Pharmacology and Therapeutics. 2010.

[15]. Baker NC, Ekins S, Williams AJ, and Tropsha A, "A bibliometric review of drug repurposing," Drug Discovery Today. 2018.

[16]. Li L et al., "Identification of type 2 diabetes subgroups through topological analysis of patient similarity," Sci. Transl. Med, 2015.

[17]. Xue H, Li J, Xie H, and Wang Y, "Review of drug repositioning approaches and resources," International Journal of Biological Sciences. 2018.

[18]. Murata M, Horiuchi E, and Kanazawa I, "Zonisamide has beneficial effects on Parkinson's disease patients," Neurosci. Res, 2001.

[19]. Nishimura Y, Tagawa M, Ito H, Tsuruma K, and Hara H, "Overcoming obstacles to drug repositioning in Japan," Front. Pharmacol, 2017.

[20]. Sirota M et al., "Discovery and preclinical validation of drug indications using compendia of public gene expression data," Sci. Transl. Med, 2011.

[21]. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, and Lu Z, "A survey of current trends in computational drug repositioning," Brief. Bioinform, 2016.

[22]. Lamb J et al., “The Connectivity Map : Using,” Science (80-.)., 2006.

[23]. Brum AM et al., “Connectivity Map-based discovery of parbendazole reveals targetable human osteogenic pathway,” Proc. Natl. Acad. Sci, 2015.

[24]. Van Noort V et al., “Novel drug candidates for the treatment of metastatic colorectal cancer through global inverse gene-expression profiling,” Cancer Res, 2014.

[25]. Dudley JT et al., “Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease,” Sci. Transl. Med, 2011.

[26]. Ferrero E and Agarwal P, “Connecting genetics and gene expression data for target prioritisation and drug repositioning,” BioData Min, 2018.

[27]. Bekris LM, Yu CE, Bird TD, and Tsuang DW, “Review article: Genetics of Alzheimer disease,” Journal of Geriatric Psychiatry and Neurology. 2010.

[28]. Riedel BC, Thompson PM, and Brinton RD, “Age, APOE and sex: Triad of risk of Alzheimer’s disease,” Journal of Steroid Biochemistry and Molecular Biology. 2016.

[29]. Beydoun MA et al., “Sex differences in the association of the apolipoprotein E epsilon 4 allele with incidence of dementia, cognitive impairment, and decline,” Neurobiol. Aging, 2012.

[30]. Neu SC et al., “Apolipoprotein E genotype and sex risk factors for Alzheimer disease: A meta-analysis,” JAMA Neurol, 2017.

[31]. Curtis AF et al., “Sex differences in the prevalence of genetic mutations in FTD and ALS,” Neurology. 2017.

[32]. Chen X, Paranjpe M, Wang R, Feng DD, Zhou Y, and for the Alzheimer’s Disease Neuroimaging Initiative, “Modeling Spatial and Temporal Patterns of APOE δ4 Mediated Glucose Uptake in Mild Cognitive Impairment and Normal Controls,” IFAC-PapersOnLine, vol. 51, no. 27, 2018.

[33]. Moreno-Grau S et al., “Exploring APOE genotype effects on Alzheimer’s disease risk and amyloid β burden in individuals with subjective cognitive decline: The FundacioACE Healthy Brain Initiative (FACEHBI) study baseline results,” Alzheimer’s and Dementia, 2017.

[34]. Manning EN et al., “APOE δ4 is associated with disproportionate progressive hippocampal atrophy in AD,” PLoS One, 2014.

[35]. Damoiseaux JS et al., “Gender Modulates the APOE 4 Effect in Healthy Older Adults: Convergent Evidence from Functional Brain Connectivity and Spinal Fluid Tau Levels,” J. Neurosci, 2012.

[36]. Liraz O, Boehm-Cagan A, and Michaelson DM, “ApoE4 induces Aβ42, tau, and neuronal pathology in the hippocampus of young targeted replacement apoE4 mice,” Mol. Neurodegener, 2013.

[37]. Paranjpe MD et al., “The effect of ApoE δ4 on longitudinal brain region-specific glucose metabolism in patients with mild cognitive impairment: a FDG-PET study,” NeuroImage Clin, vol. 22, p. 101795, 3 2019. [PubMed: 30991617]

[38]. Luchsinger JA, Reitz C, Honig LS, Tang MX, Shea S, and Mayeux R, “Aggregation of vascular risk factors and risk of incident Alzheimer disease,” Neurology, 2005.

[39]. Kivipelto M et al., “Midlife vascular risk factors and Alzheimer’s disease in later life: longitudinal, population based study.,” BMJ, 2001.

[40]. Li J et al., “Vascular risk factors promote conversion from mild cognitive impairment to Alzheimer disease,” Neurology, 2011.

[41]. Jack CR et al., “NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease,” Alzheimer’s and Dementia. 2018.

[42]. Subramanian A et al., “A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles,” Cell, 2017.

[43]. LoVerso PR, Wachter CM, and Cui F, “Cross-species transcriptomic comparison of in vitro and in vivo mammalian neural cells,” Bioinform. Biol. Insights, 2015.

[44]. Kim T et al., “In situ transcriptome characteristics are lost following culture adaptation of adult cardiac stem cells,” Sci. Rep, 2018.

[45]. Zaitseva M, Vollenhoven BJ, and Rogers PAW, “In vitro culture significantly alters gene expression profiles and reduces differences between myometrial and fibroid smooth muscle cells,” Mol. Hum. Reprod, 2006.

[46]. Yu DAK, Chen B, Charalel J, G. T, Butte A, and Sirota M, “Comprehensive transcriptomic analysis of cell lines as models of primary tumor samples across 22 tumor types,” bioRxiv, 2018.

[47]. Zhang M et al., “Drug repositioning for Alzheimer’s disease based on systematic ‘omics’ data mining,” PLoS One, 2016.

[48]. Ramos PIP et al., “An integrative, multi-omics approach towards the prioritization of Klebsiella pneumoniae drug targets,” Sci. Rep, 2018.

[49]. Buniello A et al., “The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019,” Nucleic Acids Res, 2019.

[50]. Wishart DS et al., “HMDB: The human metabolome database,” Nucleic Acids Res, 2007.

[51]. Li YH et al., “Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics,” Nucleic Acids Res, 2018.

[52]. Wishart DS, “DrugBank: a comprehensive resource for in silico drug discovery and exploration,” Nucleic Acids Res, 2006.

[53]. Gottesman O et al., “The Electronic Medical Records and Genomics (eMERGE) Network: Past, present, and future,” Genetics in Medicine. 2013.

[54]. Geifman N, Brinton RD, Kennedy RE, Schneider LS, and Butte AJ, “Evidence for benefit of statins to modify cognitive decline and risk in Alzheimer’s disease,” Alzheimer’s Res. Ther, 2017.

[55]. Zissimopoulos JM, Barthold D, Brinton RD, and Joyce G, “Sex and race differences in the association between statin use and the incidence of Alzheimer disease,” JAMA Neurol, 2017.

[56]. Paik et ah H, “Repurpose terbutaline sulfate for amyotrophic lateral sclerosis using electronic medical records,” Sci. Rep, 2015.

[57]. Metsker O, Bolgova E, Yakovlev A, Funkner A, and Kovalchuk S, “Pattern-based Mining in Electronic Health Records for Complex Clinical Process Analysis,” in Procedia Computer Science, 2017.

[58]. Delespierre T, Denormandie P, Bar-Hen A, and Josseran L, “Empirical advances with text mining of electronic health records,” BMC Med. Inform. Decis. Mak, 2017.

[59]. Ford E, Carroll JA, Smith HE, Scott D, and Cassell JA, “Extracting information from the text of electronic medical records to improve case detection: A systematic review,” J. Am. Med. Informatics Assoc, 2016.

[60]. Hripcsak G et al., “Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers,” in Studies in Health Technology and Informatics, 2015.

[61]. Malin BA and Sweeney L, “A secure protocol to distribute unlinkable health data.,” in AMIA Annual Symposium proceedings, 2005.

[62]. Dean JM, Vernon DD, Cook L, Nechodom P, Reading J, and Suruda A, “Probabilistic linkage of computerized ambulance and inpatient hospital discharge records: A potential tool for evaluation of emergency medical services,” Ann. Emerg. Med, 2001.

[63]. Milinovich A and Kattan MW, “Extracting and utilizing electronic health data from Epic for research,” Ann. Transl. Med, 2018.

[64]. Beaulieu-Jones BK and Moore JH, “MISSING DATA IMPUTATION IN THE ELECTRONIC HEALTH RECORD USING DEEPLY LEARNED AUTOENCODERS,” in Biocomputing 2017, 2017.

[65]. Wells BJ, Nowacki AS, Chagin K, and Kattan MW, “Strategies for Handling Missing Data in Electronic Health Record Derived Data,” eGEMs (Generating Evid. Methods to Improv. patient outcomes), 2013.

[66]. Bounthavong M, Watanabe JH, and Sullivan KM, “Approach to addressing missing data for electronic medical records and pharmacy claims data research,” Pharmacotherapy, 2015.

[67]. Saha B, Gupta S, and Venkatesh S, “Improved risk predictions via sparse imputation of patient conditions in electronic medical records,” in Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015, 2015.

[68]. Little RJA, “Regression with missing X’s: A review,” J. Am. Stat. Assoc, 1992.

[69]. Troyanskaya O et al., “Missing value estimation methods for DNA microarrays,” Bioinformatics, 2001.

[70]. Cowling BJ et al., “Spectral Regularization Algorithms for Learning Large Incomplete Matrices,” Euro Surveill. Bull. Eur. sur les Mal. Transm. = Eur. Commun. Dis. Bull, 2013.

[71]. Beretta L and Santaniello A, “Nearest neighbor imputation algorithms: A critical evaluation,” BMC Med. Inform. Decis. Mak, 2016.

[72]. Siderowf A and Stem MB, “Premotor Parkinson’s disease: Clinical features, detection, and prospects for treatment,” Annals of Neurology. 2008.

[73]. Ross GW et al., “Association of olfactory dysfunction with risk for future Parkinson’s disease,” Ann. Neurol, 2008.

[74]. Tolosa E, Compta Y, and Gaig C, “The premotor phase of Parkinson’s disease,” Park. Relat. Disord, 2007.

[75]. Meles SK, Teune LK, de Jong BM, Dierckx RA, and Leenders KL, “Metabolic Imaging in Parkinson Disease,” J. Nucl. Med, 2017.

[76]. Shiba M et al., “Anxiety disorders and depressive disorders preceding Parkinson’s disease: A case-control study,” Mov. Disord, 2000.

[77]. Beason-Held LLet al., “Changes in Brain Function Occur Years before the Onset of Cognitive Impairment,” J. Neurosci, 2013.

[78]. Rodriguez-Vieitez E et al., “Diverging longitudinal changes in astrocytosis and amyloid PET in autosomal dominant Alzheimer’s disease,” Brain, 2016.

[79]. Bateman RJ et al., “Clinical and Biomarker Changes in Dominantly Inherited Alzheimer’s Disease,” N. Engl. J. Med, 2012.

[80]. Quiroz et ah YT, “Association between amyloid and tau accumulation in young adults with autosomal dominant Alzheimer disease,” JAMA Neurol, 2018.

[81]. Cohen AD et al., “Basal Cerebral Metabolism May Modulate the Cognitive Effects of A in Mild Cognitive Impairment: An Example of Brain Reserve,” J. Neurosci, 2009.

[82]. Mosconi L, Berti V, Glodzik L, Pupi A, De Santi S, and De Leon MJ, “Pre-clinical detection of Alzheimer’s disease using FDG-PET, with or without amyloid imaging,” Journal of Alzheimer’s Disease. 2010.

[83]. Fox NC et al., “Presymptomatic hippocampal atrophy in Alzheimer’s disease. A longitudinal MRI study,” Brain, 1996.

[84]. Brakedal B et al., “Glitazone use associated with reduced risk of Parkinson’s disease,” Mov. Disord, 2017.

[85]. Simuni T et al., “Pioglitazone in early Parkinson’s disease: A phase 2, multicentre, double-blind, randomised trial,” Lancet Neurol, 2015.

[86]. Pihlstrøm L, Wiethoff S, and Houlden H, “Genetics of neurodegenerative diseases: an overview,” in Handbook of Clinical Neurology, 2017.

[87]. Smith KB, Kang P, and Sabbagh MN, “The effect of statins on rate of cognitive decline in mild cognitive impairment,” Alzheimer’s Dement. Transl. Res. Clin. Interv, 2017.

[88]. Feldman HH et al., “Randomized controlled trial of atorvastatin in mild to moderate Alzheimer disease: LEADe,” Neurology, 2010.

[89]. Shepherd J et al., “Pravastatin in elderly individuals at risk of vascular disease (PROSPER): A randomised controlled trial,” Lancet, 2002.

[90]. Collins R, Armitage J, Parish S, Sleight P, and Peto R, “MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: A randomised placebo-controlled trial,” Lancet, 2002.

[91]. Nance MA, “Genetics of Huntington disease,” in Handbook of Clinical Neurology, 2017.

[92]. March-Vila E et al., “On the integration of in silico drug design methods for drug repurposing,” Front. Pharmacol, 2017.

[93]. Holdgate G et al., “Biophysical methods in drug discovery from small molecule to pharmaceutical,” Methods Mol. Biol, 2013.

[94]. Talevi A, “Drug repositioning: Current approaches and their implications in the precision medicine era,” Expert Rev. Precis. Med. Drug Dev, 2018.

[95]. de Franchi E, Schalon C, Messa M, Onofri F, Benfenati F, and Rognan D, “Binding of protein kinase inhibitors to synapsin I inferred from pair-wise binding site similarity Measurements,” PLoS One, 2010.

[96]. Jalencas X and Mestres J, “Chemoisosterism in the proteome,” J. Chem. Inf. Model, 2013.

[97]. Jalencas X and Mestres J, “Identification of similar binding sites to detect distant polypharmacology,” Molecular Informatics. 2013.

[98]. Ehrt C, Brinkjost T, and Koch O, “Impact of Binding Site Comparisons on Medicinal Chemistry and Rational Molecular Design,” Journal of Medicinal Chemistry. 2016.

[99]. Fernandes HB and Raymond LA, “NMDA Receptors and Huntington’s Disease,” in Biology of the NMDA Receptor, 2009.

[100]. Hoe HS et al., “The effects of amyloid precursor protein on postsynaptic composition and activity,” J. Biol. Chem, 2009.

[101]. Song C, Zhang Y, Parsons CG, and Liu YF, “Expression of polyglutamine-expanded huntingtin induces tyrosine phosphorylation of N-methyl-D-aspartate receptors,” J. Biol. Chem, 2003.

[102]. Ehmhoefer DE, Wong BKY, and Hayden MR, “Convergent pathogenic pathways in Alzheimer’s and Huntington’s diseases: Shared targets for drug development,” Nature Reviews Drug Discovery. 2011.

[103]. Gregori-Puigjane E and Mestres J, “A Ligand-Based Approach to Mining the Chemogenomic Space of Drugs,” Comb. Chem. High Throughput Screen, 2008.

[104]. Gaulton A et al., “The ChEMBL database in 2017,” Nucleic Acids Res, 2017.

[105]. Wang Y et al., “PubChem BioAssay: 2017 update,” Nucleic Acids Res, 2017.

[106]. Stumpfe D and Bajorath J, “Exploring activity cliffs in medicinal chemistry,” Journal of Medicinal Chemistry. 2012.

[107]. Kitchen DB, Decornez H, Furr JR, and Bajorath J, “Docking and scoring in virtual screening for drug discovery: Methods and applications,” Nature Reviews Drug Discovery. 2004.

[108]. Xie H et al., “In silico drug repositioning for the treatment of Alzheimer’s disease using molecular docking and gene expression data,” RSC Adv, 2016.

[109]. Honig LS et al., “Trial of Solanezumab for Mild Dementia Due to Alzheimer’s Disease.,” N. Engl. J. Med, 2018.

[110]. Abi-Saab D, Andjelkovic M, Delmar P, Voyle N, Esau N, and Lasser RA, “THE EFFECT OF 6 MONTHS’ DOSING ON THE RATE OF AMYLOID-RELATED IMAGING ABNORMALITIES (ARIA) IN THE MARGUERITE ROAD STUDY,” Alzheimer’s Dement., 2017.

[111]. Villemagne VL et al., “A randomized, exploratory molecular imaging study targeting amyloid β with a novel 8-OH quinoline in Alzheimer’s disease: The PBT2-204 IMAGINE study,” Alzheimer’s Dement. Transl. Res. Clin. Interv, 2017.

[112]. Relkin NR et al., “A phase 3 trial of IV immunoglobulin for Alzheimer disease,” Neurology, 2017.

[113]. Doody RS et al., “A Phase 3 Trial of Semagacestat for Treatment of Alzheimer’s Disease,” N. Engl. J. Med, 2013.

[114]. Wang L, Xi Y, Sung S, and Qiao H, “RNA-seq assistant: Machine learning based methods to identify more transcriptional regulated genes,” BMC Genomics, 2018.

[115]. Smith AM et al., “Deep learning of representations for transcriptomics-based phenotype prediction,” bioRxiv, p. 574723, 3 2019.

[116]. Kadurin A, Nikolenko S, Khrabrov K, Aliper A, and Zhavoronkov A, “DruGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico,” Mol. Pharm, 2017.

[117]. Butler KT, Davies DW, Cartwright H, Isayev O, and Walsh A, “Machine learning for molecular and materials science,” Nature. 2018.

[118]. Popova M, Isayev O, and Tropsha A, “Deep reinforcement learning for de novo drug design,” Sci. Adv, 2018.

[119]. Mani S, Chen Y, Elasy T, Clayton W, and Denny J, “Type 2 diabetes risk forecasting from EMR data using machine learning.,” AMIA … Amm. Symp. proceedings. AMIA Symp, 2012.

[120]. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, and Buchman TG, “An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU,” Crit. Care Med, 2018.

[121]. SHAMEER K et al., “PREDICTIVE MODELING OF HOSPITAL READMISSION RATES USING ELECTRONIC MEDICAL RECORD-WIDE MACHINE LEARNING: A CASE-STUDY USING MOUNT SINAI HEART FAILURE COHORT,” 2016.

[122]. Bakkar N et al., “Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis,” Acta Nenropathol, 2018.

[123]. Napolitano F et al., “Drug repositioning: A machine-learning approach through data integration,” J. Cheminform, 2013.

[124]. Kim DH, Wit H, and Thurston M, “Artificial intelligence in the diagnosis of Parkinson’s disease from ioflupane-123 single-photon emission computed tomography dopamine transporter scans using transfer learning,” Nucl. Med. Commun, 2018.

[125]. Blahuta J and Soukup T, “Ultrasound medical image recognition with artificial intelligence for Parkinson’s disease classification,” … 2012 Proc …,2012.

[126]. Zhao Y et al., “Exploration of machine learning techniques in predicting multiple sclerosis disease course,” PLoS One, 2017.

[127]. Yoo Y et al., “Deep learning of brain lesion patterns for predicting future disease activity in patients with early symptoms of multiple sclerosis,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016.

[128]. Moradi E, Pepe A, Gaser C, Huttunen H, and Tohka J, “Machine learning framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects,” Neuroimage, 2015.

[129]. Zhang D and Shen D, “Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease,” Neuroimage, 2012.

[130]. Li H, Habes M, and Fan Y, “A DEEP LEARNING PROGNOSTIC MODEL FOR EARLY PREDICTION OF ALZHEIMER’S DISEASE BASED ON HIPPOCAMPAL MRIDATA,” Alzheimer’s Dement, 2018.

[131]. Wilding JL and Bodmer WF, “Cancer cell lines for drug discovery and development,” Cancer Research. 2014.

[132]. Dawson TM, Golde TE, and Lagier-Tourenne C, “Animal models of neurodegenerative diseases,” Nature Neuroscience. 2018.

[133]. Ameen-Ali KE, Wharton SB, Simpson JE, Heath PR, Sharp P, and Berwick J, “Review: Neuropathology and behavioural features of transgenic murine models of Alzheimer’s disease,” Neuropathology and Applied Neurobiology. 2017.

[134]. Kang S, Lee YH, and Lee JE, “Metabolism-centric overview of the pathogenesis of Alzheimer’s disease,” Yonsei Medical Journal. 2017.

[135]. Frega M, Tedesco M, Massobrio P, Pesce M, and Martinoia S, “Network dynamics of 3D engineered neuronal cultures: A new experimental model for in-vitro electrophysiology,” Sci. Rep, 2014.

[136]. Krencik Ret al., “Systematic Three-Dimensional Coculture Rapidly Recapitulates Interactions between Human Neurons and Astrocytes,” Stem Cell Reports, 2017.

[137]. Gaudry KS, “Evergreening: a common practice to protect new drugs,” Nat. Biotechnol, 2011.

[138]. Shineman DWet al., “Overcoming obstacles to repurposing for neurodegenerative disease,” Ann. Clin. Transl. Neurol, 2014.

[139]. King A, “The search for better animal models of Alzheimer’s disease,” Nature, 2018.

[140]. Mann KM et al., “Independent effects of APOE on cholesterol metabolism and brain Aβ levels in an Alzheimer disease mouse model,” Hum. Mol Genet, 2004.

[141]. Cheng-Hathaway PJ et al., “The Trem2 R47H variant confers loss-of-function-like phenotypes in Alzheimer’s disease,” Mol. Neurodegener, 2018.

[142]. Mullane K and Williams M, “Preclinical Models of Alzheimer’s Disease: Relevance and Translational Validity,” Curr. Protoc. Pharmacol, 2019.

[143]. Lo E, “Sharing data from clinical trials: Where we are and what lies ahead,” BMJ, 2013.

Outstanding Questions

- How can we create animal and cell models that better phenocopy human neurodegenerative disease to aid in validation of drug repurposing hits?
- How can we use available molecular and clinical data to better model neurodegenerative diseases with long disease courses that require years or decades of follow-up?
- How can we develop and apply more sophisticated AI and ML algorithms to generate computational drug predictions by integrating disparate data modalities?
- How can we increase non-profit investment in drug repurposing?

Highlights

- Computational drug repurposing has the potential to drastically reduce the cost and development time for therapies for neurodegenerative disease.
- Advent of artificial intelligence (AI) and machine learning (ML) algorithms will facilitate the integration of several modalities of data to advance computational drug repurposing for neurodegenerative disease.
- The creation of large scale transcriptomic and EMR databases provides a novel opportunity for computational drug repurposing.
- Computational drug repurposing for neurodegenerative disease is uniquely challenging due to the lack of efficacious validation methods and long and heterogeneous disease course.

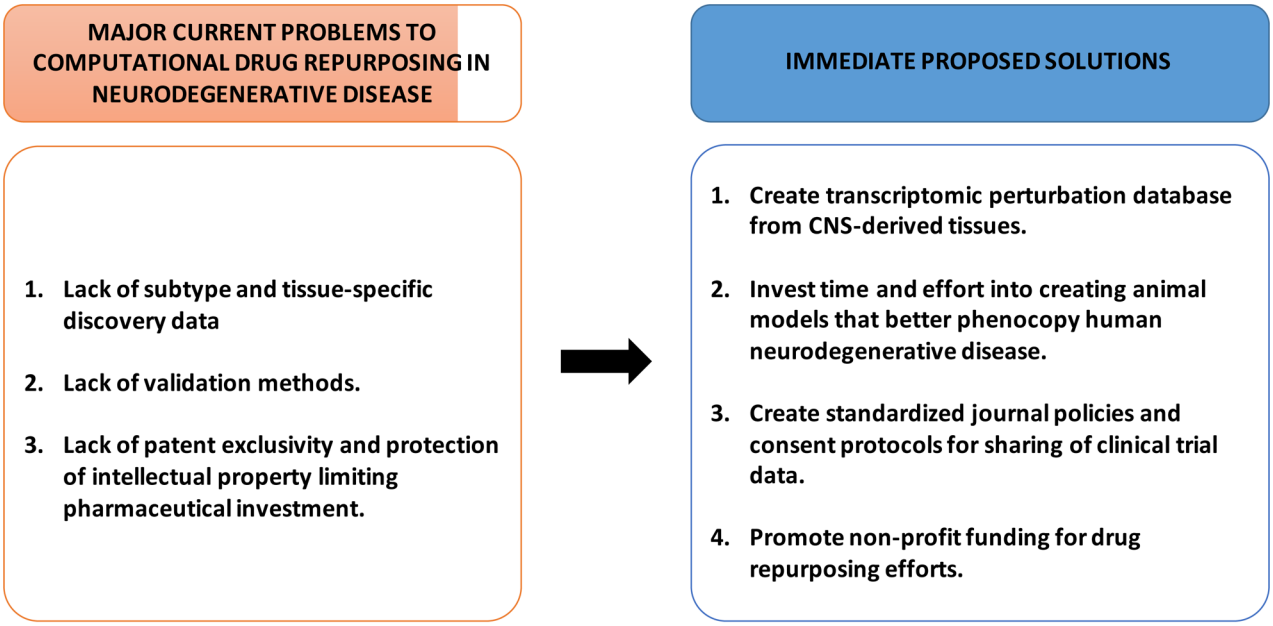

Figure 1:

Major limitations and future solutions to computational drug repurposing in neurodegenerative disease. The current major limitations to computational drug repurposing in neurodegenerative disease are listed on the left panel. To address these limitations, we propose several solutions that are presented in the right panel.

Table 1:

Current Strategies for Computational Drug Repurposing

	Description	Advantages	Disadvantages	Examples of Preclinical Successes
Molecular	Compare drug gene expression signatures pre- and post-drug treatment to disease gene expression signatures in order to predict drugs that may reverse disease gene signatures.	• Does not require *a priori* identification of target molecule. • Can integrate genetic, epigenetic and transcriptomic data	• Limited by the availability of drug and disease molecular profiling data. • Requires creation of CNS-derived transcriptomic perturbation dataset.	• Parbendazole for osteoporosis [23] • Cimetidine for lung adenocarcinoma [20] • Citalopram for metastatic colon cancer[24] • Topiramate for inflammatory bowel disease [25]
Clinical	Leverage large-scale health data such as electronic medical records and patient medication histories to identify drugs effective for indications other than the primary use.	• Large amounts of health data can be obtained from Electronic Medical Record (EMR), insurance claims data, clinical trial data, health registries, health surveys, personal genome testing companies • With sufficient sample size, precision medicine approaches can be utilized. • Disparate data sources such as EMR and claims data can be standardized using the OMOP formalism.	• Clinical data must be converted into a structured database before analysis. • EMR data is often messy and incomplete • Patients must be longitudinally profiled in case of neurodegenerative disease. • Health outcomes are difficult to track for neurodegenerative courses with long disease courses. • Clinical data should be paired with genetic data to enable genetic-subtype specific drug repurposing.	• Statin therapy for AD [54], [55] • Terbutaline sulfate for ALS [56] • Glitazone for PD [84]
Biophysical	Leverage the biochemical properties of drugs such as binding affinity or biophysical properties such as 3D conformation to achieve drug-target predictions.	• Useful in neurodegenerative diseases with known targets. • Computationally efficient mechanism to screen thousands of drug molecules with high-throughput techniques.	• Biophysical methods require crystallographic data of target and drug molecules. • Require *a priori*. identification of target molecules.	• Droperidol for AD [108]