Review

Changing Trends in Computational Drug Repositioning

Jaswanth K. Yella [1], Suryanarayana Yaddanapudi [1], Yunguan Wang [1] and Anil G. Jegga [1,2,3,*]

[1] Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center,
    240 Albert Sabin Way MLC 7024, Cincinnati, OH 45229, USA; Jaswanth.Yella@cchmc.org (J.K.Y.);
    Suryanarayana.Yaddanapudi@cchmc.org (S.Y.); Yunguan.Wang@cchmc.org (Y.W.)
[2] Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA
[3] Department of Computer Science, University of Cincinnati College of Engineering,
    Cincinnati, OH 45219, USA
[*] Correspondence: anil.jegga@cchmc.org; Tel.: +1-(513)-636-0261; Fax: +1-(513)-636-2056

check for updates

Abstract: Efforts to maximize the indications potential and revenue from drugs that are already marketed are largely motivated by what Sir James Black, a Nobel Prize-winning pharmacologist advocated—"The most fruitful basis for the discovery of a new drug is to start with an old drug". However, rational design of drug mixtures poses formidable challenges because of the lack of or limited information about in vivo cell regulation, mechanisms of genetic pathway activation, and in vivo pathway interactions. Hence, most of the successfully repositioned drugs are the result of "serendipity", discovered during late phase clinical studies of unexpected but beneficial findings. The connections between drug candidates and their potential adverse drug reactions or new applications are often difficult to foresee because the underlying mechanism associating them is largely unknown, complex, or dispersed and buried in silos of information. Discovery of such multi-domain *pharmacomodules*—pharmacologically relevant sub-networks of biomolecules and/or pathways—from collection of databases by independent/simultaneous mining of multiple datasets is an active area of research. Here, while presenting some of the promising bioinformatics approaches and pipelines, we summarize and discuss the current and evolving landscape of computational drug repositioning.

Keywords: computational drug repositioning; drug repositioning; drug repurposing; machine learning; deep learning; crowdsourcing; open innovation; drug discovery

1. Introduction

The path to new drug discovery has always been a road full of twists and turns. De novo drug discovery in particular is an expensive, time-consuming, and high risk process. For instance, the total average cost of developing a new drug, as per an estimate, ranges from $2 billion to $3 billion and it takes at least 13–15 years to bring a drug to the market—starting from initial discovery to the approval stage [1]. Further, the process suffers from a high rate of attrition. About 10% of the drugs that enter into clinical trials get approved by regulatory agencies [2]. The remaining 90% of the drugs fail due to inefficacy or high toxicity due to the limited predictive value of preclinical studies [3]. Nearly 62% of the compounds fail in Phase II and approximately 45% attrition occurs in Phase III [4]. These attritions are due to insufficient R&D productivity in identifying the drug response on the target due to the limited availability of preclinical disease models which has raised concerns in the pharmaceutical industry [5]. Despite rapid technological advances and exponential increases in pharmaceutical R&D investments, the number of newly approved drugs continues to be the same [6]. To overcome these

challenges and to potentially bypass this productivity gap, more and more companies are resorting to "drug repositioning" or "drug repurposing" (sometimes also referred to as drug reprofiling, drug retasking, or therapeutic switching) or simply identifying and developing new therapeutic uses for existing or abandoned pharmacotherapies [7]. The premise is that since most approved compounds have known bioavailability and safety profiles, proven formulation and manufacturing routes, and reasonably characterized pharmacology, repositioned drugs can enter clinical phases more rapidly and at a lower cost than novel compounds. Further, the 90% therapeutic development failure rate means there are many existing, partially developed therapeutic candidates that could be re-visited, explored further, and potentially repurposed for a new disease, common or rare. It is therefore not surprising that in recent years, of the new drugs that reach their first markets, repositioned drugs have taken up to a percentage of ~30%! For instance, of the 113 new drugs and biologics approved or launched in 2017, only seven were first-in-class agents (an approved and launched first drug with a novel mechanism of action) while 36 were repositioned drugs [8]. As per an estimate, this bypassing can potentially make a drug available for use in patients within 3–12 years with a total estimated cost of $40–80 million [9,10].

Table 1. Examples of repositioned drugs (adapted in part from [11], this list is neither extensive nor exhaustive).

Drug	Original Indication	New Indication
Allopurinol	Cancer	Gout
Amantadine	Influenza	Parkinson's disease
Amphotericin	Antifungal	Leishmaniasis
Arsenic	Syphilis	Leukemia
Aspirin	Inflammation, pain	Antiplatelet
Atomexetine	Depressive disorder	ADHD
Bimatoprost	Glaucoma	Promoting eyelash growth
Bromocriptine	Parkinson's disease	Diabetes mellitus
Bupropion	Depression	Smoking cessation
Colchicine	Gout	Recurrent pericarditis
Colesevelam	Hyperlipidemia	Type 2 diabetes mellitus
Dapsone	Leprosy	Malaria
Disulfiram	Alcoholism	Melanoma
Doxepin	Depressive disorder	Antipruritic
Eflornithine	Depression	ADHD
Finasteride	Benign prostatic hyperplasia	Male pattern baldness
Gabapentin	Epilepsy	Neuropathic pain
Gemcitabine	Antiviral	Cancer
Lomitapide	Lipidemia	Familial hypercholesterolemia
Methotrexate	Cancer	Psoriasis, rheumatoid arthritis
Miltefosine	Cancer	Visceral leishmaniasis
Minoxidil	Hypertension	Hair loss
Naltrexone	Opioid addiction	Alcohol withdrawal
Naproxen	Inflammation, pain	Alzheimer's disease
Nortriptyline	Depression	Neuropathic pain
Premetrexed	Mesothelioma	Lung cancer
Propranolol	Hypertension	Migraine prophylaxis
Raloxifene	Contraceptive	Osteoporosis
Sildenafil	Angina	Erectile dysfunction; pulmonary hypertension
Thalidomide	Morning sickness	Leprosy; multiple myeloma
Tretinoin	Acne	Leukemia
Zidovudine	Cancer	HIV/AIDS
Zileuton	Asthma	Acne

Most of the successful cases of drug repurposing have been serendipitous discoveries rather than systematic, hypothesis-driven outcomes. These include the accidental discovery of thalidomide as an agent for leprosy or the more notable example of sildenafil, an angina medication developed in 1989 subsequently marketed as Viagra®, a blockbuster drug to treat erectile dysfunction [12] (see Table 1 for

additional examples of drug repositioning). De novo drug therapies for more than 8000 orphan or rare diseases are impossible to develop with the current R&D costs, however, drug repositioning with its premise of discovering hidden connections or building connections between a drug and disease hold promise for orphan disease therapy [13]. Further, revisiting the approved drugs for identifying new indications helps the pharmaceutical companies to extend the patent life of drugs, through application to adjacent diseases and also helps the company to protect the IP against competitors [14].

In-silico methods like data-mining, machine learning, and network-based approaches, offer an unprecedented opportunity to predict all possible drug repositioning candidates using available diverse and heterogeneous data sources from genomics and biomedical domains [15]. Indeed, predictive models have been built using these methods exploiting existing data such as protein targets, chemical structure, or phenotypic information such as profiles of side-effect, gene expression, etc. While the advances in computational sciences bring the possibility of applying novel algorithms and approaches to systems biology data, these datasets themselves have triggered fundamental research on more complex problems [16]. As a result of this hybrid approach of utilizing computational methods and experimental screenings, various modalities of drug repositioning methods have emerged. Computational drug repositioning methods focus on shared characteristics between two drugs and depending on what kind of drug discovery (drug-based or disease-based) [17], the methods can be classified in to target-based, expression-based, knowledge-based, chemical structure-based, pathway-based and mechanism of action-based [18]. In this review article, we briefly outline the recent progress in computational methods and strategies applied on the drug-disease data for drug repositioning investigations.

2. Approaches

In silico drug repurposing challenges that are drug-centric (i.e., discovering new indications for existing drugs) or disease-centric (i.e., identifying an effective drug as a potential treatment for disease) have the common challenge of either assessing the similarity or connections between drugs or between diseases [19]. Jin and Wong [18] reviewed a variety of approaches used as a basis for computational drug repurposing. These can be broadly categorized as knowledge-based and signature-based approaches.

2.1. Knowledge-Based Drug Repurposing

This repurposing method utilizes the available information on drug such as drug-targets, chemical structures, adverse effects, pathways etc. and builds computational models to predict unknown mechanisms, targets or new bio-markers for diseases [20–24]. In pathway-based approach, signaling pathways, metabolic pathways and protein-interaction networks data are used to compute the similarity or connections between drug and disease. The processed omics data, for example, from human patients or animal models of disease are used to reconstruct disease-specific pathways that can serve as key targets for novel therapeutic discovery or for repositioned drugs [25–30]. Target mechanism-based approaches on the other hand take into account known mechanism of action and target role : Here, the data available on signaling pathways, protein interactions and omics data are integrated to identify the potential mechanism of action (MoA) of drugs [31–34]. This in turn can enable find better and even specific drug targets and also for discover of an alternate medication for any disease.

2.2. Signature-Based Drug Repurposing

This method makes use of gene expression signatures by comparing drug gene expression profiles and disease gene expression profiles and is frequently referred to as 'signature reversion' method [35]. Gene expression based methods are effective in constructing a detailed map of connections between diseases and drug actions [36–40].

Table 2. Drug and Disease Centric Database Resources.

Database	Type	Description	URL	Ref.
ADReCS	Drug	System Toxicology and in silico drug safety evaluation. Contains 137,619 Drug-ADR pairs	http://bioinf.xmu.edu.cn/ADReCS/	[41]
ChEMBL	Drug	Database of bioactive drug-like small molecules and abstracted bioactivities	https://www.ebi.ac.uk/chembl	[42]
ChemSpider	Drug	Database of 64 million chemical structures	http://www.chemspider.com/	[43]
Clue (L1000 Platform)	Drug	Dataset of transcriptional responses of human cells to chemical and genetic perturbation. 1.2 Million L1000 profiles and tools for their analysis.	https://clue.io/	[44]
Comparative Toxicogenomics Database	Drug	Associations of Drug-Gene, Gene-Disease, Drug-Disease and gene-gene	http://ctdbase.org/	[45]
DailyMED	Drug	Catalogue of drug listings/drug label information	https://dailymed.nlm.nih.gov/dailymed/	[46]
DGIdb	Drug	Drug-gene annotations, interactions and potential drug ability database	http://dgidb.org/	[47]
DrugBank	Drug	Contains 11,000 drug entries and each entry contains more than 200 data fields of chemical information and drug targets.	https://www.drugbank.ca/	[48]
DrugCentral	Drug	Information on active ingredients chemical entities, pharmaceutical products, drug mode of action, indications, pharmacologic action	http://drugcentral.org/	[49]
e-Drug3D	Drug	e-Drug3D offers a facility to explore FDA approved drugs and active metabolites	http://chemoinfo.ipmc.cnrs.fr/MOLDB/index.html	[50]
Genomics of Drug Sensitivity in Cancer (GDSC)	Drug	Screenings of >1000 genetically characterized human cancer cell lines with a wide range of anti-cancer therapeutics	http://www.cancerrxgene.org/	[51]
Inxight Drugs	Drug	A comprehensive portal for drug development information from NCATS	https://drugs.ncats.io/ginas/app	
Open Targets Platform	Drug	comprehensive and robust data integration for access to and visualization of potential drug targets associated with disease	https://www.targetvalidation.org	[52]
PharmGKB	Drug	Curated dataset of genetic variation on drug response	https://www.pharmgkb.org/	[53]

Table 2. *Cont.*

Database	Type	Description	URL	Ref.
pkCSM	Drug	Small-molecule pharmacokinetic (ADMET) properties prediction using SMILE data	http://biosig.unimelb.edu.au/pkcsm/prediction	[54]
Project Achilles	Drug	A genome-wide catalog of tumor dependencies, to identify vulnerabilities associated with genetic and epigenetic alterations	https://portals.broadinstitute.org/achilles	[55]
Promiscuous	Drug	Database contains three different types of entities: drugs, proteins and side-effects as well as relations between them	http://bioinformatics.charite.de/promiscuous/	[56]
PubChem	Drug	PubChem contains more than 90 million compounds chemical information along with their bio activities, gene and protein targets	http://pubchem.ncbi.nlm.nih.gov/	[57]
SIDER	Drug	Information on marketed medicines and their recorded adverse drug reactions	http://sideeffects.embl.de/	[58]
STITCH	Drug	68,000 chemicals, interactions and over 1.5 million proteins in 373 species	http://stitch.embl.de/	[59]
SuperPred	Drug	A prediction webserver for ATC code and target prediction of compounds	http://prediction.charite.de/	[60]
Therapeutic Target Database (TTD)	Drug	Dataset of known and explored therapeutic protein and nucleic acid targets, the targeted disease, pathway information and the corresponding drugs directed at each of these target	http://bidd.nus.edu.sg/group/cjttd/	[61]
Toxin and Toxin-Target Database (T3DB)	Drug	A database of 3673 toxins described by 41,733 synonyms, including pollutants, pesticides, drugs, and food toxins, which are linked to 2087 corresponding toxin target records	http://www.t3db.ca/	[62]
Human Protein Atlas	Disease and Drug	Consists of three separate parts; the Tissue Atlas showing the distribution of the proteins across all major tissues and organs in the human body, the Cell Atlas showing the subcellular localization of proteins in single cells, and finally the Pathology Atlas showing the impact of protein levels for survival of patients with cancer.	https://www.proteinatlas.org/	[63]
KEGG Medicus	Disease and Drug	Collection of databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances	http://www.genome.jp/kegg/disease/ http://www.kegg.jp/ http://www.genome.jp/kegg/drug/	[64]

Table 2. *Cont.*

Database	Type	Description	URL	Ref.
PsychEncode	Disease		https://www.synapse.org//#!Synapse:syn4921369/wiki/235539	[65]
Allen Brain Atlas	Disease	Gene expression maps for mouse and human brain	http://www.brain-map.org/	[66]
ArrayExpress	Disease	Micro array gene expression data at EBI	https://www.ebi.ac.uk/arrayexpress	[67]
CCLE	Disease	Database of mRNA expression and mutation data over 1100 cancer cell lines	https://portals.broadinstitute.org/ccle	[68]
COSMIC	Disease	Catalogue of somatic mutations in human cancer	http://cancer.sanger.ac.uk/cosmic	[69]
dbGAP	Disease	Catalogue of somatic mutations causing cancer	http://www.ncbi.nlm.nih.gov/gap	[70]
dbSNP	Disease	Database of single nucleotide polymorphisms	https://www.ncbi.nlm.nih.gov/snp	[71]
dbVar	Disease	Public archives for genomic structural variation	https://www.ncbi.nlm.nih.gov/dbvar	[72]
DisGeNET	Disease	Database on human disease-associated genes and variants	http://www.disgenet.org/	[73]
ENCODE	Disease	Database of comprehensive parts list of functional elements in human genome	https://genome.ucsc.edu/ENCODE/	[20]
Genomics Data Commons	Disease	Harmonized Cancer Datasets with 40 cancer mutated gene projects, 22,147 Genes and 3 million mutations	https://gdc.cancer.gov/	[74]
GEO	Disease	High throughput gene expression datasets	http://www.ncbi.nlm.nih.gov/geo	[75]
GTex	Disease	Catalog of genetic variations and their influence on gene expressions	https://www.gtexportal.org/home/	[76]
Human Proteome Map	Disease	Interactive resource with massive peptide sequencing results	http://www.humanproteomemap.org/	[77]
ICGC	Disease	Dataset with more than 17,000 cancer donors spanning 76 projects and 21 tumor sites	http://icgc.org/	[78]
IGSR	Disease	1000 genome project data usability and extension	http://www.internationalgenome.org/	[79]
Orphadata	Disease	Rare diseases, drugs and associated genes	http://www.orphadata.org/cgi-bin/index.php/	[80]
Roadmap Epigenomics	Disease	Epigenomic maps for stem cells and primary ex vivo tissues selected to represent the normal counterparts of tissues and organ systems frequently involved in human disease	http://www.roadmapepigenomics.org/	[81]
STRING	Disease	Protein-Protein interaction, analysis, and networks	https://string-db.org/cgi/input.pl	[82]

Connectivity Map (CMap) [83,84], NCBI's Gene Expression (GEO) [75], and the relatively recent LINCS datasets [44] are also extensively explored in drug repositioning studies. Recent technical and technological advancements in molecular biology and exponential growth of biomedical data while presenting challenges have also opened up an array of opportunities to develop and apply novel and powerful computational approaches that can enable informed drug repositioning. The free availability of data repositories are further directing and catalyzing these efforts. In Table 2 we present some of the widely used open source drug- and disease-centric and related databases. These include, for instance, databases that provide information on the known targets, mechanism of action, gene expression, clinical status, ADMET properties, signaling pathways and disease-centric database which has omics data (transcriptomic, proteomic, genetic characteristics of diseases).

2.3. In Silico Methods for Drug Repositioning

In the following sections, we present an overview of some of the in silico methods—current and emerging—used for facilitating drug repositioning candidate discovery.

2.3.1. Machine Learning

Any machine learning workflow typically comprises of 4 steps: data pre-processing, feature extraction, model fitting and evaluation [85]. PREDICT, is a similarity based machine learning framework, integrating drug-drug similarity (based on drug-protein interactions, sequence and gene-ontology) and disease-disease similarity (disease-phenotype and human phenotype ontology) where the authors have used them as features applying logistic regression to predict similar drugs for similar diseases and they achieved AUC = 0.9 in predicting drug indications [86]. SPACE, another similarity-based method predicts anatomical therapeutic chemical classification of drugs by integrating multiple data sources using Logistic Regression [87]. Likewise, several such similarity based methods have been reported for predicting novel drug indications [88–90].

Deep learning, a large class of machine learning-based models composed of multiple processing layers representing data with a high level of abstraction are now being explored computational biology field for a wide-variety of applications including drug discovery [91,92]. The principal difference between conventional "shallow" learning (neural network with one or two hidden layers) and deep learning is that while the former does not deal with raw data and requires a feature extraction step to be performed before the learning process, the latter not only discovers intricate structure in large data sets but by using the backpropagation algorithm allows changing the internal parameters incrementally to compute the representation in each layer from the representation in the previous layer [92]. Deep learning-based approaches have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and are currently being explored in biomedical and genomic domains. Aliper and Plis, for example, used deep learning with gene expression data to learn drug therapeutic categories and found that deep neural networks surpassed SVM after 10 fold cross validation suggesting a working proof for applying deep learning for drug discovery and development [93]. Interestingly, Zhao and Cheong, compared deep neural networks (DNN) approach with SVM-based approach to predict psychiatric drug indications based on the expression profiles of drugs and reported that [37]. While more studies are needed to understand if DNN-based approaches indeed have the claimed benefits, there have been additional reports suggesting that deep learning-based approaches perform better than traditional machine learning algorithms in toxicity prediction by enabling multi-task learning [94,95].

2.3.2. Network Models

Network-based approaches have been extensively exploited in computational drug repositioning for identifying novel drug targets, interactions, and indications [96]. Typically, in these models, the nodes in the networks represent either drug, disease, or gene products and edges represent the interactions or relationships between them. These networks are either knowledge-based or

computationally inferred using multiple data resources and have various representations such as drug-drug, drug-target, drug-disease, disease-disease, disease-gene, disease-drug, protein-protein interactions, and transcriptional networks [97]. Cheng and Liu computed similarities—drug-based, target-based, and network-based—to predict drug-target interaction in a bi-partite network and found that network based inference method performed best with an average ROC AUC of 0.96 [21]. Similar homogenous or bipartite network models have been incorporated using phenotype data such as side-effect [98–100], transcriptional [101–103], drug-disease [104,105] and signaling pathway data [25].

Integrating heterogeneous data also provides diverse information and has the potential to unveil hidden or unknown drug-disease relationships based on the guilt-by-association principle. Most of the similarity-based methods are either drug-centric or disease-centric networks, with relatively few approaches that built a drug-disease heterogeneous network using compendia of gene annotations and network clustering to identify drug repositioning candidates [105,106]. Luo and Zhao, built a similar network-based framework using heterogeneous data through a network diffusion process and used the diffusion distributions to derive the prediction scores of drug-target interactions [107]. Recently, Himmelstein et al. integrated data from 29 public resources to identify dug repositioning candidates and predicted the probability of repositioning for 209,168 drug-disease pairs [108].

2.3.3. Mining Electronic Health Records for Drug Repurposing

Electronic health records (EHR) of the patients which provide medications details along with patient history can also be mined to identify drug repositioning candidates. Applying natural language processing on EHRs, for instance, reveals post-market, additional adverse drug events which are not found in clinical trials [109]. These side-effects information can be potentially used for drug-repositioning and validation [23]. Mining EHR records for example helped in identifying that metformin, a most commonly prescribed medication for type II diabetes, can also be repurposed for cancer treatment [110]. The relevance and accuracy of the model's prediction needs to be assessed in discovering a drug whose indications are unknown. The validity of novel drug prediction can be evaluated by comparing the predicted targets in ClinicalTrials.gov, PubMed abstracts or EHR records. The performance of the model can be evaluated by computing area under the ROC curve (AUC ROC) and Precision Recall (PR) curve. Sensitivity is a metric to measure the proportion of true positive identified correctly and Specificity is the proportion of negatives correctly identified as negatives. Due to the large unannotated drug-indication pairs as false positives, the sensitivity and specificity estimates are poor and creates substantial imbalance of true positives and true negatives. In a recent review, Brown and Patel suggest that using sensitivity-validation alone is ideal since it does not need the true negatives. The authors further suggest that investigators should test their model performance with cross-validation to prevent over-fitting and weak predictive performance [111].

2.4. Open Innovation—Crowd Sourcing

Crowd sourcing is a collaborative approach of delegating tasks to the crowd where the variety of expertise available generates new insights or hypothesis with the available data. This paradigm has been taken advantage in a multitude areas from diverse domains including health care and genomics. The open source drug discovery process enables faster translation of research to results with a clear definition on specific problem, task decomposition and immediate feedback loop [112–115]. Pharmaceutical companies, due to the limitations in R&D business model and man power often are focused on specific diseases which may or may not include rare and neglected diseases. Hence, few pharmaceutical and non-profit companies have used crowdsourcing platforms and embraced a wide-variety of innovative solutions [116,117] directing towards discussing the scientific enigmas. Several open innovation platforms have been established in order to build industry-academia partnerships and to explore science and business opportunities with mutual benefit (Table 3).

Table 3. Open innovation research resources.

Name	Description	URL
Centers for Therapeutic Innovation (CTI)	Collaborative research platform for clinical applications and drug discovery [118]	https://www.pfizercti.com
CREEDS	Crowd-extracted expression of differential signatures [119]	http://amp.pharm.mssm.edu/CREEDS
Grants4Leads	Financial support for exploration of new approaches in infectious diseases [120]	https://www.grants4leads.com/
Kaggle	Data scientists and statisticians competition platform with few bioinformatics challenges [117,121]	http://www.kaggle.com/
Open Innovation Drug Discovery	Academic and Industry researchers open collaboration platform for drug discovery [122]	https://openinnovation.lilly.com/dd/
Sage Bionetworks	Bioinformatics and data science challenge platform building prognostic models for breast cancer [123]	http://sagebionetworks.org/
TopCoder	Machine learning engineers, programmers and data scientists challenge platform [116]	http://www.topcoder.com

National Center for Advancing Translational Sciences (NCATS)—NIH-Academia-Industry Partnerships Initiative

The National Institutes of Health (NIH), as part of the new therapeutic uses program, launched (i) NCATS' NIH-Industry Partnerships initiative in 2012 to foster collaboration between pharmaceutical companies and the biomedical research community; and (ii) bench-to clinical repurposing initiative to test the utility of crowdsourcing efforts or computational approaches for drug repurposing.

The focus of the *match-making* NIH-industry partnerships projects is to match researchers with *open assets* from pharmaceutical assets to fuel and accelerate drug repurposing candidate discovery. Through this initiative, NCATS supports and advances research on a wide range of common and rare (including neglected) diseases. Current industry partners in this initiative include: AstraZeneca, AbbVie, Bristol-Myers Squibb, Eli Lilly, GlaxoSmithKline, Janssen Pharmaceuticals, MedImmune, Mereo BioPharma, Pfizer, and Sanofi. The participating companies make a number of partially developed assets available to academic researchers to crowdsource repurposing ideas. Projects using most of these assets can go directly into Phase II clinical trials, while some may require additional pre-clinical investigations or a Phase I clinical trial (e.g., testing in target populations to determine dosing, assess safety and tolerability).

Through the bench-to-clinic repurposing program, NCATS supports pre-clinical studies, clinical feasibility studies or proof-of-concept clinical trials to assess the utility of computational approaches or crowdsourcing efforts in discovering drug repurposing candidates. Table 4 lists the new therapeutic uses projects funded by NIH-NCATS through these two programs (additional details can be found at https://ncats.nih.gov/ntu/projects).

Table 4. NIH-NCATS funded new therapeutic uses projects (2013–2018).

Project/Study Title	Year	NCATS Program	Condition
The Efficacy and Safety of a Selective Estrogen Receptor Beta Agonist (LY500307)	2013	NIH-Industry Partnership	Schizophrenia
Fyn Inhibition by AZD0530 for Alzheimer's Disease	2013	NIH-Industry Partnership	Alzheimer's disease
Medication Development of a Novel Therapeutic for Smoking Cessation	2013	NIH-Industry Partnership	Cigarette smoking
A Novel Compound for Alcoholism Treatment: A Translational Strategy	2013	NIH-Industry Partnership	Alcoholism
Partnering to Treat an Orphan Disease: Duchenne Muscular Dystrophy	2013	NIH-Industry Partnership	Duchenne muscular dystrophy
Reuse of ZD4054 for Patients with Symptomatic Peripheral Artery Disease	2013	NIH-Industry Partnership	Peripheral artery disease
Therapeutic Strategy for Lymphangioleiomyomatosis	2013	NIH-Industry Partnership	Lymphangioleiomyomatosis
Therapeutic Strategy to Slow Progression of Calcific Aortic Valve Stenosis	2013	NIH-Industry Partnership	Calcific aortic valve stenosis
Translational Neuroscience Optimization of GlyT1 Inhibitor	2013	NIH-Industry Partnership	Schizophrenia
Anti-inflammatory Small Drug as Adjunctive Therapy to Improve Glucometabolic Variables in Obese, Insulin-Resistant Type 2 Diabetic Patients	2015	NIH-Industry Partnership	Insulin-resistant type 2 diabetes
Evaluation of AZD9291 in Glioblastoma Patients with Activated EGFR	2015	NIH-Industry Partnership	Glioblastoma
Evaluation of a Cathepsin S Inhibitor as a Potential Drug for Chagas Disease	2015	NIH-Industry Partnership	Chagas disease
Wee1 and HDAC Inhibition in Relapsed/Refractory AML	2015	NIH-Industry Partnership	Relapsed/refractory AML
Anti-Virulence Drug Repurposing Using Structural Systems Pharmacology	2016	Bench-to-Clinic	Bacterial virulence
CXCR2 Antagonism in the Immunometabolic Regulation of Type 2 Diabetes	2016	Bench-to-Clinic	Type 2 diabetes
Drug Repositioning in Diabetic Nephropathy	2016	Bench-to-Clinic	Diabetic nephropathy
Ketorolac and Related NSAIDs for Targeting Rho-Family GTPases in Ovarian Cancer	2016	Bench-to-Clinic	Ovarian cancer
Network-Driven Drug Repurposing Approaches to Treat Coronary Artery Disease	2016	Bench-to-Clinic	Coronary artery disease
Pre-Clinical Evaluation of a Neutrophil Elastase Inhibitor for the Treatment of Inflammatory Bowel Disease	2016	Bench-to-Clinic	Inflammatory bowel disease
Quantum Model Repurposing of Cethromycin for Liver Stage Malaria	2016	Bench-to-Clinic	Liver-stage malaria
Repurposing Lesogaberan for the Treatment of Type 1 Diabetes	2016	Bench-to-Clinic	Type 1 diabetes
Repurposing Misoprostol for Clostridium Difficile Colitis as Identified by PheWAS	2016	Bench-to-Clinic	Clostridium difficile colitis

Table 4. *Cont.*

Project/Study Title	Year	NCATS Program	Condition
Repurposing Pyronaridine as a Treatment for the Ebola Virus	2016	Bench-to-Clinic	Ebola virus
Therapeutic Repurposing of Benserazide for Colon Cancer	2016	Bench-to-Clinic	Colon cancer
Computational Repurposing of Chemotherapies for Pulmonary Hypertension	2017	Bench-to-Clinic	Pulmonary hypertension
Pre-Clinical Evaluation of Vorinostat in Alopecia Areata	2017	Bench-to-Clinic	Alopecia areata
Pre-Clinical Testing of a Novel Therapeutic for Nonalcoholic Steatohepatitis	2017	Bench-to-Clinic	Nonalcoholic steatohepatitis
Repurposing Pyronaridine as a Treatment for Chagas Disease	2017	Bench-to-Clinic	Chagas disease
Single-Cell-Driven Drug Repositioning Approaches to Target Inflammation in Atherosclerosis	2017	Bench-to-Clinic	Atherosclerosis
Impact of SAR152954 on Prenatal Alcohol Exposure-Induced Neurobehavioral Deficits	2017	Bench-to-Clinic	Neurobehavioral deficits
An Endoplasmic Reticulum Calcium Stabilizer for the Treatment of Wolfram Syndrome	2017	Bench-to-Clinic	Wolfram syndrome
Utilization of Phenotypic Precision Medicine to Identify Optimal Drug Combinations for the Treatment of Hepatocellular Carcinoma	2017	Bench-to-Clinic	Hepatocellular carcinoma
Targeting Glucose Metabolism for the Treatment of Hepatocellular Carcinoma	2017	Bench-to-Clinic	Hepatocellular carcinoma
Application of a Repurposed FDA Approved Drug as a Local Osteogenic Agent	2017	Bench-to-Clinic	To induce local osteogenesis
Repurposing Misoprostol to Prevent Recurrence of Clostridium Difficile Infection	2018	Bench-to-Clinic	Recurrent Clostridium difficile
AZD9668: A First in Class Disease Modifying Therapy to Treat Alpha-1 Antitrypsin Deficiency, a Genetically Linked Orphan Disease	2018	NIH-Industry Partnership	Alpha-1 antitrypsin deficiency
AZD9668 and Neutrophil Elastase Inhibition to Prevent Graft-versus-Host Disease	2018	NIH-Industry Partnership	Graft-versus-host disease
Use of the Src Family Kinase Inhibitor Saracatinib in the Treatment of Pulmonary Fibrosis	2018	NIH-Industry Partnership	Pulmonary fibrosis

2.5. Open Source Software

The open source movement has created a substantial value in pursuing towards "*state-of-the-art*" research over the last decade with the help of reusable and generic software libraries for data processing [124,125]. Jupyter Notebook for instance is the modern data analysis tool for reproducible computational research that supports open source languages like Python, Julia, C++, R and several other languages and provides rich features for interactive computing, visualization, and documentation [126]. Structured data tools like Scikit-learn [127], R-Programming, Orange [128] and Weka [129] are useful for mining, analysis, learning and statistical computing. For the high dimensional un-structured data such as images, text, or audio outputs, deep learning tools like TensorFlow [130], Keras [131], PyTorch [132], CNTK [133], and Matlab [134] that take advantage of multi-GPU accelerated training are increasingly used. Gephi [135] and Cytoscape [136] are other popular tools used primarily for bimolecular interaction networks, omics-data integration, clustering and visualization. In Table 5, we summarize few such used tools used in computational drug discovery and repositioning.

Table 5. Web-tools and open source kits.

Tool	Description	URL	Ref
Clue	Tools for perturbagens (small molecules or genes) query, L1000 cohorts, and gene expression heatmap visualization	https://clue.io	[44]
Clue Repurposing Tool	Interactive application to access approved and pre-clinical drug annotations	https://clue.io/repurposing	[137]
COGENA	Analysis, visualizing and clustering tool for gene expression profiles	https://github.com/zhilongjia/cogena	[138]
DeepChem	Deep learning toolkit for drug discovery and cheminformatics	https://deepchem.io/	[139]
DR.PRODIS	Prediction of drug-protein interactions, side effects	http://cssb.biology.gatech.edu/repurpose	[140]
e-LEA3D	Collection of tools related to computer-aided drug design	http://chemoinfo.ipmc.cnrs.fr/	[141]
Frog2	Chemo-informatics toolkit for small compound 3D generation from 1D/2D input	http://bioserv.rpbs.univ-paris-diderot.fr/services/Frog2/	[142]
GIFT	Infer chemogenomic features from drug-target interactions.	http://bioinfo.au.tsinghua.edu.cn/software/GIFT/	[143]
GoPredict	Drug target prioritization tool for breast and ovarian cancer	http://csblcanges.fimm.fi/GOPredict/	[144]
JOELib/JOELib2	Toolkit to interconvert chemical file formats, descriptor calculation classes, and SMARTS substructure search	http://www.ra.cs.uni-tuebingen.de/software/joelib/introduction.html	[145]
ksRepo	Drug repositioning tool that utilizes gene expression drug datasets from different platforms	https://github.com/adam-sam-brown/ksRepo	[101]
L1000CDS	L1000 dataset based gene expression signature search engine	http://amp.pharm.mssm.edu/L1000CDS2/#/index	[146]
MANTRA	Prediction and analysis of mechanism of action of drugs for drug repositioning	http://mantra.tigem.it/	[147]
NFFinder	Tool to discover multiple drugs with similar drugs based on up/down regulated genes	http://nffinder.cnb.csic.es/	[102]
Open babel	Open source chemistry toolbox	http://openbabel.org/wiki/Main_Page	[148]
Open PHACTS	European funded initiative to bring together industry and academic partners for semantic integration of pharmacological data using an RDF data model	http://www.openphacts.org	[149]

3. Discussion

Drug repositioning acts as a viable strategy for a cost-effective de novo drug discovery. Although in silico methods have proven to be successful in addressing the problem of repurposing, some

challenges continue to be addressed. One of the principal issues is the missing drug-disease indication data. Marking the missing indications as true negatives or ignoring them from training can potentially compromise the predictive power of the computational model for drug repurposing candidate discovery. Second, the lack of a true gold standard dataset for drug repositioning makes it difficult for in silico methods to evaluate results. As a result, common performance metrics such as sensitivity, specificity, and precision are used to assess the utility of computational drug repurposing algorithms. Third, existing computational methods tend to be predominantly one-sided (e.g., drug-centric or disease-centric). However, the integration of multi-omic data with similarity measures have been shown to have better predictive performance with identification of novel therapeutic compounds [105,108].

The sea of biomedical information (see Table 2), in which small molecule and gene/protein structural, functional and process knowledge—both in normal and disease states—is embedded consists of unstructured free-text as in publications and structured or semi-structured relational databases. Transforming information from these silos into actionable knowledge is facilitated by establishing connectivity among the subsets taken from these multiple heterogeneous and diverse domains. For example, a *pharmacomodule* consisting of a group of genes, biological processes, pathways, phenotypes, small molecules (approved drugs or investigational compounds), and a group of drug-induced or related adverse events forms a meaningful multi-domain module when the interdependency among most of the pairs of subsets are supported by scientific evidence (literature or databases). These *pharmacomodules* can potentially take us closer to answering the *how* question about the underlying a hypothetical mechanism of action or phenomena. An informed answer to the *how* question holds the premise to generate better and informed drug repositioning hypotheses. Growing scientific evidence [7] suggests that any compound found to be safe in humans is likely to have multiple therapeutic uses. However, almost all successful drug reposition crossovers so far have been the result of either accidental occurrences or informed guesses. Given that this "back-to-basics" approach for repositioning is growing in popularity [8], there is an urgent need for more efficient and systematic computational approaches to first systematize the available genomic and pharmacological databases for representation and knowledge discovery and then use these databases and pattern discovery tools to identify the potential new uses for existing drugs. What is needed clearly is a paradigm shift in the approaches—genomic, biopharmacological, and computational—for a more informed systematic drug rediscovery ("systematic serendipity") taking into account all of the data resources. Originally coined by Eugene Garfield, "systematic serendipity" refers to the organized process of discovering previously unknown scientific relations using citation databases, leading to better possibilities for a collaboration of human serendipity with computer supported knowledge discovery [150].

The credibility of published research will improve the discoveries in science if the provided compendium has an evidence for the accuracy and reproducibility of the results. Reproducibility particularly is a major issue especially when scientific papers publish unexpected, positive results and other researchers or an independent research group is unable to replicate the same results even after using same or similar methods as reported by the original study [151,152]. It has been estimated that the irreproducible research costs up to $28 billion per year [153]! Providing the code and data used to obtain the claimed and reported results is always a better strategy than mere describing them in natural language in the paper and can be eventually an incremental step towards a better science [154–156]. The recent Findability, Accessibility, Interoperability, Reusability (FAIR) data principles go beyond the mere reuse of data by individuals but rather enhance the ability of machines to support and find and use the data automatically. These include any efforts that support discovery and reproducibility through good data management practices such as good data management, maintenance of the data flow, and sharing relevant tools or pipelines used in the research [157]. The recent Datasets2Tools project is in compliance with these principles and enables users to search for contributed canned analyses, datasets and tools [158]. Computational science research can be replicated effectively using tools like code version control software like Github [159] and transferable computational environments

like Docker [160]. Over the past few years, the reproducibility issue is being taken seriously and many journals insist on providing code and data when submitting the paper.

In summary, emerging and advanced novel computational methods and crowdsourcing-based approaches that enable the joint analysis of genomic, biomedical and pharmacological data hold the premise to facilitate informed, efficient, and systematic drug repositioning. Whether this premise expedites drug development pipelines and how much of it translates into novel therapeutic discovery and impacts public health, especially catering to unmet needs (e.g., rare and neglected diseases), positively remains to be seen.

References

1. Scannell, J.W.; Blanckley, A.; Boldon, H.; Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **2012**, *11*, 191–200. [PubMed]
2. Akhondzadeh, S. The Importance of Clinical Trials in Drug Development. *Avicenna J. Med. Biotechnol.* **2016**, *8*, 151. [PubMed]
3. Plenge, R.M.; Scolnick, E.M.; Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **2013**, *12*, 581–594. [CrossRef] [PubMed]
4. Kola, I.; Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **2004**, *3*, 1–5. [CrossRef] [PubMed]
5. Paul, S.M.; Mytelka, D.S.; Dunwiddie, C.T.; Persinger, C.C.; Munos, B.H.; Lindborg, S.R.; Schacht, A.L. How to improve RD productivity: The pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **2010**, *9*, 203–214. [CrossRef] [PubMed]
6. Booth, B.; Zemmel, R. Opinion/Outlook: Prospects for productivity. *Nat. Rev. Drug Discov.* **2004**, *3*, 451–456. [CrossRef] [PubMed]
7. Ashburn, T.T.; Thor, K.B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **2004**, *3*, 673–683. [CrossRef] [PubMed]
8. Graul, A.I.; Cruces, E.; Stringer, M. The year's new drugs & biologics, 2013: Part I. *Drugs of Today* **2014**, *50*, 51–100. [PubMed]
9. Hurle, M.R.; Yang, L.; Xie, Q.; Rajpal, D.K.; Sanseau, P.; Agarwal, P. Computational drug repositioning: From data to therapeutics. *Clin. Pharmacol. Ther.* **2013**, *93*, 335–341. [CrossRef] [PubMed]
10. Papapetropoulos, A.; Szabo, C. Inventing new therapies without reinventing the wheel: The power of drug repurposing. *Br. J. Pharmacol.* **2018**, 2016–2018. [CrossRef] [PubMed]
11. Padhy, B.M.; Gupta, Y.K. Drug repositioning: Re-investigating existing drugs for new therapeutic indications. *J. Postgrad. Med.* **2011**, *57*, 153–160. [CrossRef] [PubMed]
12. Nosengo, N. Can you teach old drugs new tricks? *Nature* **2016**, *534*, 314–316. [CrossRef] [PubMed]
13. Sardana, D.; Zhu, C.; Zhang, M.; Gudivada, R.C.; Yang, L.; Jegga, A.G. Drug repositioning for orphan diseases. *Brief. Bioinform.* **2011**, *12*, 346–356. [CrossRef] [PubMed]
14. Cha, Y.; Erez, T.; Reynolds, I.J.; Kumar, D.; Ross, J.; Koytiger, G.; Kusko, R.; Zeskind, B.; Risso, S.; Kagan, E.; et al. Drug repurposing from the perspective of pharmaceutical companies. *Br. J. Pharmacol.* **2017**. [CrossRef] [PubMed]
15. Li, J.; Zheng, S.; Chen, B.; Butte, A.J.; Swamidass, S.J.; Lu, Z. A survey of current trends in computational drug repositioning. *Brief. Bioinform.* **2016**, *17*, 2–12. [CrossRef] [PubMed]
16. Prathipati, P.; Mizuguchi, K. Systems Biology Approaches to a Rational Drug Discovery Paradigm. *Curr. Top. Med. Chem.* **2015**, *16*, 1009–1025. [CrossRef]
17. Li, Y.Y.; Jones, S.J.M. Drug repositioning for personalized medicine. *Genome Med.* **2012**, *4*, 27. [PubMed]
18. Jin, G.; Wong, S.T.C. Toward better drug repositioning: Prioritizing and integrating existing methods into efficient pipelines. *Drug Discov. Today* **2014**, *19*, 637–644. [CrossRef] [PubMed]

19. Liu, Z.; Fang, H.; Reagan, K.; Xu, X.; Mendrick, D.L.; Slikker, W.; Tong, W. In silico drug repositioning: What we need to know. *Drug Discov. Today* **2013**, *18*, 110–115. [CrossRef] [PubMed]

20. Emig, D.; Ivliev, A.; Pustovalova, O.; Lancashire, L.; Bureeva, S.; Nikolsky, Y.; Bessarabova, M. Drug Target Prediction and Repositioning Using an Integrated Network-Based Approach. *PLoS ONE* **2013**, *8*. [CrossRef] [PubMed]

21. Cheng, F.; Liu, C.; Jiang, J.; Lu, W.; Li, W.; Liu, G.; Zhou, W.; Huang, J.; Tang, Y. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.* **2012**, *8*. [CrossRef] [PubMed]

22. Zhao, S.; Li, S. Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS ONE* **2010**, *5*. [CrossRef] [PubMed]

23. Yang, L.; Agarwal, P. Systematic drug repositioning based on clinical side-effects. *PLoS ONE* **2011**, *6*. [CrossRef] [PubMed]

24. Kinnings, S.L.; Liu, N.; Buchmeier, N.; Tonge, P.J.; Xie, L.; Bourne, P.E. Drug discovery using chemical systems biology: Repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol.* **2009**, *5*. [CrossRef] [PubMed]

25. Jadamba, E.; Shin, M. A Systematic Framework for Drug Repositioning from Integrated Omics and Drug Phenotype Profiles Using Pathway-Drug Network. *BioMed Res. Int.* **2016**, *2016*. [CrossRef] [PubMed]

26. Li, J.; Lu, Z. Pathway-based drug repositioning using causal inference. *BMC Bioinform.* **2013**, *14*. [CrossRef] [PubMed]

27. Kotelnikova, E.; Yuryev, A.; Mazo, I.; Daraselia, N. Computational approaches for drug repositioning and combination therapy design. *J. Bioinform. Comput. Biol.* **2010**, *8*, 593–606. [CrossRef] [PubMed]

28. Cramer, P.; Cirrito, J.; Wesson, D. ApoE-Directed Therapeutics Rapidly Clear beta-Amyloid and Reverse Deficits in AD Mouse Models. *Science* **2012**, *335*, 1503–1506. [CrossRef] [PubMed]

29. Sivachenko, A.; Kalinin, A.; Yuryev, A. Pathway Analysis for Design of Promiscuous Drugs and Selective Drug Mixtures. *Curr. Drug Discov. Technol.* **2006**, *3*, 269–277. [CrossRef] [PubMed]

30. Strittmatter, W.J. Old drug, new hope for Alzheimer's disease. *Science* **2012**, *335*, 1447–1448. [CrossRef] [PubMed]

31. Jin, G.; Zhao, H.; Zhou, X.; Wong, S.T.C. An enhanced Petri-Net model to predict synergistic effects of pairwise drug combinations from gene microarray data. *Bioinformatics* **2011**, *27*. [CrossRef] [PubMed]

32. Jin, G.; Fu, C.; Zhao, H.; Cui, K.; Chang, J.; Wong, S.T.C. A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy. *Cancer Res.* **2012**, *72*, 33–44. [CrossRef] [PubMed]

33. Iskar, M.; Zeller, G.; Blattmann, P.; Campillos, M.; Kuhn, M.; Kaminska, K.H.; Runz, H.; Gavin, A.C.; Pepperkok, R.; Van Noort, V.; et al. Characterization of drug-induced transcriptional modules: Towards drug repositioning and functional understanding. *Mol. Syst. Biol.* **2013**, *9*. [CrossRef] [PubMed]

34. Gaiteri, C.; Ding, Y.; French, B.; Tseng, G.C.; Sibille, E. Beyond modules and hubs: The potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav.* **2014**, *13*, 13–24. [CrossRef] [PubMed]

35. Iorio, F.; Rittman, T.; Ge, H.; Menden, M.; Saez-Rodriguez, J. Transcriptional data: A new gateway to drug repositioning? *Drug Discov. Today* **2013**, *18*, 350–357. [CrossRef] [PubMed]

36. Sirota, M.; Dudley, J.T.; Kim, J.; Chiang, A.P.; Morgan, A.A.; Sweet-Cordero, A.; Sage, J.; Butte, A.J. Discovery and preclinical validation of drug indications using compendia of public gene expression data (Science Translational Medicine (2011) 3, (102er)). *Sci. Transl. Med.* **2011**, *3*, 96ra77. [CrossRef] [PubMed]

37. Zhao, K.; So, H.-C. A machine learning approach to drug repositioning based on drug expression profiles: Applications in psychiatry. *arXiv* **2017**, arXiv:1706.03014.

38. Wang, Y.; Yella, J.; Chen, J.; McCormack, F.X.; Madala, S.K.; Jegga, A.G. Unsupervised gene expression analyses identify IPF-severity correlated signatures, associated genes and biomarkers. *BMC Pulm. Med.* **2017**, *17*. [CrossRef] [PubMed]

39. Claerhout, S.; Lim, J.Y.; Choi, W.; Park, Y.-Y.; Kim, K.; Kim, S.-B.; Lee, J.-S.; Mills, G.B.; Cho, J.Y. Gene Expression Signature Analysis Identifies Vorinostat as a Candidate Therapy for Gastric Cancer. *PLoS ONE* **2011**, *6*, e24662. [CrossRef] [PubMed]

40. Chang, M.; Smith, S.; Thorpe, A.; Barratt, M.J.; Karim, F. Evaluation of phenoxybenzamine in the CFA model of pain following gene expression studies and connectivity mapping. *Mol. Pain* **2010**, *6*. [CrossRef] [PubMed]

41. Cai, M.C.; Xu, Q.; Pan, Y.J.; Pan, W.; Ji, N.; Li, Y.B.; Liu, H.J.J.K.; Ji, Z.L. ADReCS: An ontology database for aiding standardization and hierarchical Classification of adverse drug reaction terms. *Nucleic Acids Res.* **2015**, *43*, D907–D913. [CrossRef] [PubMed]

42. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*. [CrossRef] [PubMed]

43. Williams, A.J. Internet-based tools for communication and collaboration in chemistry. *Drug Discov. Today* **2008**, *13*, 502–506. [CrossRef] [PubMed]

44. Subramanian, A.; Narayan, R.; Corsello, S.M.; Peck, D.D.; Natoli, T.E.; Lu, X.; Gould, J.; Davis, J.F.; Tubelli, A.A.; Asiedu, J.K.; et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**, *171*, 1437–1452. [CrossRef] [PubMed]

45. Mattingly, C.J.; Colby, G.T.; Forrest, J.N.; Boyer, J.L. The Comparative Toxicogenomics Database (CTD). *Environ. Health Perspect.* **2003**, *111*, 793. [CrossRef] [PubMed]

46. National Institutes of Health: Health & Human Services DailyMed. Available online: http://dailymed.nlm.nih.gov/dailymed/index.cfm (accessed on 25 April 2018).

47. Griffith, M.; Griffith, O.L.; Coffman, A.C.; Weible, J.V.; Mcmichael, J.F.; Spies, N.C.; Koval, J.; Das, I.; Callaway, M.B.; Eldred, J.M.; et al. DGIdb: Mining the druggable genome. *Nat. Methods* **2013**, *10*, 1209–1210. [CrossRef] [PubMed]

48. DrugBank DrugBank. Available online: http://www.drugbank.ca (accessed on 22 April 2018).

49. Ursu, O.; Holmes, J.; Knockel, J.; Bologa, C.G.; Yang, J.J.; Mathias, S.L.; Nelson, S.J.; Oprea, T.I. DrugCentral: Online drug compendium. *Nucleic Acids Res.* **2017**, *45*, D932–D939. [CrossRef] [PubMed]

50. Pihan, E.; Colliandre, L.; Guichou, J.F.; Douguet, D. E-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics* **2012**, *28*, 1540–1541. [CrossRef] [PubMed]

51. Yang, W.; Soares, J.; Greninger, P.; Edelman, E.J.; Lightfoot, H.; Forbes, S.; Bindal, N.; Beare, D.; Smith, J.A.; Thompson, I.R.; et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **2013**, *41*. [CrossRef] [PubMed]

52. Koscielny, G.; An, P.; Carvalho-Silva, D.; Cham, J.A.; Fumis, L.; Gasparyan, R.; Hasan, S.; Karamanis, N.; Maguire, M.; Papa, E.; et al. Open Targets: A platform for therapeutic target identification and Validation. *Nucleic Acids Res.* **2017**, *45*, D985–D994. [CrossRef] [PubMed]

53. Hewett, M.; Oliver, D.E.; Rubin, D.L.; Easton, K.L.; Stuart, J.M.; Altman, R.B.; Klein, T.E. PharmGKB: The pharmacogenetics knowledge base. *Nucleic Acids Res.* **2002**, *30*, 163–165. [CrossRef] [PubMed]

54. Pires, D.E.V.; Blundell, T.L.; Ascher, D.B. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* **2015**, *58*, 4066–4072. [CrossRef] [PubMed]

55. Cowley, G.S.; Weir, B.A.; Vazquez, F.; Tamayo, P.; Scott, J.A.; Rusin, S.; East-Seletsky, A.; Ali, L.D.; Gerath, W.F.J.; Pantel, S.E.; et al. Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci. Data* **2014**, *1*. [CrossRef] [PubMed]

56. Von Eichborn, J.; Murgueitio, M.S.; Dunkel, M.; Koerner, S.; Bourne, P.E.; Preissner, R. PROMISCUOUS: A database for network-based drug-repositioning. *Nucleic Acids Res.* **2011**, *39*. [CrossRef] [PubMed]

57. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [CrossRef] [PubMed]

58. Kuhn, M.; Letunic, I.; Jensen, L.J.; Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **2016**, *44*, D1075–D1079. [CrossRef] [PubMed]

59. Kuhn, M.; von Mering, C.; Campillos, M.; Jensen, L.J.; Bork, P. STITCH: Interaction networks of chemicals and proteins. *Nucleic Acids Res.* **2008**, *36*. [CrossRef] [PubMed]

60. Dunkel, M.; Günther, S.; Ahmed, J.; Wittig, B.; Preissner, R. SuperPred: Drug classification and target prediction. *Nucleic Acids Res.* **2008**, *36*. [CrossRef] [PubMed]

61. Chen, X.; Ji, Z.L.; Chen, Y.Z. TTD: Therapeutic Target Database. *Nucleic Acids Res.* **2002**, *30*, 412–415. [CrossRef] [PubMed]

62. Wishart, D.; Arndt, D.; Pon, A.; Sajed, T.; Guo, A.C.; Djoumbou, Y.; Knox, C.; Wilson, M.; Liang, Y.; Grant, J.; et al. T3DB: The toxic exposome database. *Nucleic Acids Res.* **2015**, *43*, D928–D934. [CrossRef] [PubMed]

63. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [CrossRef] [PubMed]

64. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **1999**, *27*, 29–34. [CrossRef] [PubMed]

65. Akbarian, S.; Liu, C.; Knowles, J.A.; Vaccarino, F.M.; Farnham, P.J.; Crawford, G.E.; Jaffe, A.E.; Pinto, D.; Dracheva, S.; Geschwind, D.H.; et al. The PsychENCODE project. *Nat. Neurosci.* **2015**, *18*, 1707–1712. [CrossRef] [PubMed]

66. Sunkin, S.M.; Ng, L.; Lau, C.; Dolbeare, T.; Gilbert, T.L.; Thompson, C.L.; Hawrylycz, M.; Dang, C. Allen Brain Atlas: An integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.* **2013**, *41*, D996–D1008. [CrossRef] [PubMed]

67. Parkinson, H. ArrayExpress–a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **2004**, *33*, D553–D555. [CrossRef] [PubMed]

68. Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A.A; Kim, S.; Wilson, C.J.; Lehár, J.; Kryukov, G.V; Sonkin, D.; et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity Supp. *Nature* **2012**, *483*, 603–607. [CrossRef] [PubMed]

69. Forbes, S.A.; Beare, D.; Gunasekaran, P.; Leung, K.; Bindal, N.; Boutselakis, H.; Ding, M.; Bamford, S.; Cole, C.; Ward, S.; et al. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **2015**, *43*, D805–D811. [CrossRef] [PubMed]

70. Mailman, M.D.; Feolo, M.; Jin, Y.; Kimura, M.; Tryka, K.; Bagoutdinov, R.; Hao, L.; Kiang, A.; Paschall, J.; Phan, L.; et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **2007**, *39*, 1181–1186. [CrossRef] [PubMed]

71. Sherry, S.T. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, *29*, 308–311. [CrossRef] [PubMed]

72. Lappalainen, I.; Lopez, J.; Skipper, L.; Hefferon, T.; Spalding, J.D.; Garner, J.; Chen, C.; Maguire, M.; Corbett, M.; Zhou, G.; et al. DbVar and DGVa: Public archives for genomic structural variation. *Nucleic Acids Res.* **2013**, *41*. [CrossRef] [PubMed]

73. Piñero, J.; Queralt-Rosinach, N.; Bravo, À.; Deu-Pons, J.; Bauer-Mehren, A.; Baron, M.; Sanz, F.; Furlong, L.I. DisGeNET: A discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015**, *2015*. [CrossRef] [PubMed]

74. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A.; Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **2016**, *375*, 1109–1112. [CrossRef] [PubMed]

75. Barrett, T. NCBI GEO: Mining millions of expression profiles–database and tools. *Nucleic Acids Res.* **2004**, *33*, D562–D566. [CrossRef] [PubMed]

76. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [CrossRef] [PubMed]

77. Kim, M.-S.; Pinto, S.M.; Getnet, D.; Nirujogi, R.S.; Manda, S.S.; Chaerkady, R.; Madugundu, A.K.; Kelkar, D.S.; Isserlin, R.; Jain, S.; et al. A draft map of the human proteome. *Nature* **2014**, *509*, 575–581. [CrossRef] [PubMed]

78. Cancer, T. International Cancer Genome Consortium. *Cancer* **2011**, *2011*, 1–20. [CrossRef]

79. Clarke, L.; Fairley, S.; Zheng-Bradley, X.; Streeter, I.; Perry, E.; Lowy, E.; Tassé, A.M.; Flicek, P. The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res.* **2017**, *45*, D854–D859. [CrossRef] [PubMed]

80. Aymé, S.; Schmidtke, J. Networking for rare diseases: A necessity for Europe. *Bundesgesundheitsblatt Gesundheitsforschunq Gesundheitsschutz* **2007**, *50*, 1477–1483. [CrossRef] [PubMed]

81. Bernstein, B.E.; Stamatoyannopoulos, J.A.; Costello, J.F.; Ren, B.; Milosavljevic, A.; Meissner, A.; Kellis, M.; Marra, M.A.; Beaudet, A.L.; Ecker, J.R.; et al. The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* **2010**, *28*, 1045–1048. [CrossRef] [PubMed]

82. Szklarczyk, D.; Franceschini, A.; Kuhn, M.; Simonovic, M.; Roth, A.; Minguez, P.; Doerks, T.; Stark, M.; Muller, J.; Bork, P.; et al. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **2011**, *39*. [CrossRef] [PubMed]

83. Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.P.; Subramanian, A.; Ross, K.N.; et al. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**, *313*, 1929–1935. [CrossRef] [PubMed]

84. Musa, A.; Ghoraie, L.S.; Zhang, S.-D.; Galzko, G.; Yli-Harja, O.; Dehmer, M.; Haibe-Kains, B.; Emmert-Streib, F. A review of connectivity map and computational approaches in pharmacogenomics. *Brief. Bioinform.* **2017**, bbw112. [CrossRef] [PubMed]

85. Angermueller, C.; Pärnamaa, T.; Parts, L.; Oliver, S. Deep Learning for Computational Biology. *Mol. Syst. Biol.* **2016**, 878. [CrossRef] [PubMed]

86. Gottlieb, A.; Stein, G.Y.; Ruppin, E.; Sharan, R. PREDICT: A method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **2011**, *7*. [CrossRef] [PubMed]

87. Liu, Z.; Guo, F.; Gu, J.; Wang, Y.; Li, Y.; Wang, D.; Lu, L.; Li, D.; He, F. Similarity-based prediction for Anatomical Therapeutic Chemical classification of drugs by integrating multiple data sources. *Bioinformatics* **2015**, *31*, 1788–1795. [CrossRef] [PubMed]

88. Li, J.; Lu, Z. A new method for computational drug repositioning using drug pairwise similarity. In Proceedings of the 2012 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2012, Philadelphia, PA, USA, 4–7 October 2012; pp. 453–456.

89. Luo, H.; Wang, J.; Li, M.; Luo, J.; Peng, X.; Wu, F.X.; Pan, Y. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* **2016**, *32*, 2664–2671. [CrossRef] [PubMed]

90. Oh, M.; Ahn, J.; Yoon, Y. A network-based classification model for deriving novel drug-disease associations and assessing their molecular actions. *PLoS ONE* **2014**, *9*. [CrossRef] [PubMed]

91. Mamoshina, P.; Vieira, A.; Putin, E.; Zhavoronkov, A. Applications of Deep Learning in Biomedicine. *Mol. Pharm.* **2016**, *13*, 1445–1454. [CrossRef] [PubMed]

92. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

93. Aliper, A.; Plis, S.; Artemov, A.; Ulloa, A.; Mamoshina, P.; Zhavoronkov, A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* **2016**, *13*, 2524–2530. [CrossRef] [PubMed]

94. Unterthiner, T.; Mayr, A.; Klambauer, G.; Hochreiter, S. Toxicity Prediction using Deep Learning. *Front. Environ. Sci.* **2015**, *3*, 10. [CrossRef]

95. Ramsundar, B.; Kearnes, S.; Riley, P.; Webster, D.; Konerding, D.; Pande, V. Massively Multitask Networks for Drug Discovery. *arXiv* **2015**, arXiv:1502.02072.

96. Schadt, E.E. Molecular networks as sensors and drivers of common human diseases. *Nature* **2009**, *461*, 218–223. [CrossRef] [PubMed]

97. Azuaje, F. Drug interaction networks: An introduction to translational and clinical applications. *Cardiovasc. Res.* **2013**, *97*, 631–641. [CrossRef] [PubMed]

98. Ye, H.; Liu, Q.; Wei, J. Construction of drug network based on side effects and its application for drug repositioning. *PLoS ONE* **2014**, *9*. [CrossRef] [PubMed]

99. Mizutani, S.; Pauwels, E.; Stoven, V.; Goto, S.; Yamanishi, Y. Relating drug-protein interaction network with drug side effects. *Bioinformatics* **2012**, *28*. [CrossRef] [PubMed]

100. Campillos, M.; Kuhn, M.; Gavin, A.-C.; Jensen, L.J.; Bork, P. Drug target identification using side-effect similarity. *Science* **2008**, *321*, 263–266. [CrossRef] [PubMed]

101. Brown, A.S.; Kong, S.W.; Kohane, I.S.; Patel, C.J. ksRepo: A generalized platform for computational drug repositioning. *BMC Bioinform.* **2016**, *17*. [CrossRef] [PubMed]

102. Setoain, J.; Franch, M.; Martínez, M.; Tabas-Madrid, D.; Sorzano, C.O.S.; Bakker, A.; Gonzalez-Couto, E.; Elvira, J.; Pascual-Montano, A. NFFinder: An online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning. *Nucleic Acids Res.* **2015**, *43*, W193–W199. [CrossRef] [PubMed]

103. Chang, R.; Shoemaker, R.; Wang, W. A novel knowledge-driven systems biology approach for phenotype prediction upon genetic intervention. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2011**, *8*, 1170–1182. [CrossRef] [PubMed]

104. Chen, H.; Zhang, H.; Zhang, Z.; Cao, Y.; Tang, W. Network-based inference methods for drug repositioning. *Comput. Math. Methods Med.* **2015**, *2015*. [CrossRef] [PubMed]

105. Wu, C.; Gudivada, R.C.; Aronow, B.J.; Jegga, A.G. Computational drug repositioning through heterogeneous network clustering. *BMC Syst. Biol.* **2013**, *7*, 1–9. [CrossRef] [PubMed]

106. Chen, J.; Bardes, E.E.; Aronow, B.J.; Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **2009**, *37*. [CrossRef] [PubMed]

107. Luo, Y.; Zhao, X.; Zhou, J.; Yang, J.; Zhang, Y.; Kuang, W.; Peng, J.; Chen, L.; Zeng, J. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* **2017**, *8*. [CrossRef] [PubMed]

108. Himmelstein, D.S.; Lizee, A.; Hessler, C.; Brueggeman, L.; Chen, S.L.; Hadley, D.; Green, A.; Khankhanian, P.; Baranzini, S.E. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife* **2017**, *6*, 1–35. [CrossRef] [PubMed]

109. Luo, Y.; Thompson, W.K.; Herr, T.M.; Zeng, Z.; Berendsen, M.A.; Jonnalagadda, S.R.; Carson, M.B.; Starren, J. Natural Language Processing for EHR-Based Pharmacovigilance: A Structured Review. *Drug Saf.* **2017**, *40*, 1075–1089. [CrossRef] [PubMed]

110. Xu, H.; Aldrich, M.C.; Chen, Q.; Liu, H.; Peterson, N.B.; Dai, Q.; Levy, M.; Shah, A.; Han, X.; Ruan, X.; et al. Validating drug repurposing signals using electronic health records: A case study of metformin associated with reduced cancer mortality. *J. Am. Med. Inform. Assoc.* **2014**, 1–10. [CrossRef] [PubMed]

111. Brown, A.S.; Patel, C.J. A review of validation strategies for computational drug repositioning. *Brief. Bioinform.* **2018**, *19*, 174–177. [CrossRef] [PubMed]

112. Lee, W.H. Open access target validation is a more efficient way to accelerate drug discovery. *PLoS Biol.* **2015**, *13*, 1–9. [CrossRef] [PubMed]

113. Carter, A.J.; Donner, A.; Lee, W.H.; Bountra, C. Establishing a reliable framework for harnessing the creative power of the scientific crowd. *PLoS Biol.* **2017**, *15*. [CrossRef] [PubMed]

114. Rijnders, T.W.; Tzalis, D.; Jaroch, S. The European lead factory—An experiment in colla- borative drug discovery. *J. Med. Dev. Sci.* **2015**, *1*, 20–33.

115. Munos, B. Can open-source drug RD repower pharmaceutical innovation? *Clin. Pharmacol. Ther.* **2010**, *87*, 534–536. [CrossRef] [PubMed]

116. Lakhani, K.R.; Boudreau, K.J.; Loh, P.R.; Backstrom, L.; Baldwin, C.; Lonstein, E.; Lydon, M.; MacCormack, A.; Arnaout, R.A.; Guinan, E.C. Prize-based contests can provide solutions to computational biology problems. *Nat. Biotechnol.* **2013**, *31*, 108–111. [CrossRef] [PubMed]

117. Bentzien, J.; Muegge, I.; Hamner, B.; Thompson, D.C. Crowd computing: Using competitive dynamics to develop and refine highly predictive models. *Drug Discov. Today* **2013**, *18*, 472–478. [CrossRef] [PubMed]

118. Patel, A.C.; Coyle, A.J. Building a new biomedical ecosystem: Pfizer's centers for therapeutic innovation. *Clin. Pharmacol. Ther.* **2013**, *94*, 314–316. [CrossRef] [PubMed]

119. Wang, Z.; Monteiro, C.D.; Jagodnik, K.M.; Fernandez, N.F.; Gundersen, G.W.; Rouillard, A.D.; Jenkins, S.L.; Feldmann, A.S.; Hu, K.S.; McDermott, M.G.; et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat. Commun.* **2016**, *7*. [CrossRef] [PubMed]

120. Dorsch, H.; Jurock, A.E.; Schoepe, S.; Lessl, M.; Asadullah, K. Grants4Targets: An open innovation initiative to foster drug discovery collaborations. *Nat. Rev. Drug Discov.* **2014**, *14*, 74. [CrossRef] [PubMed]

121. Markoff, J. Scientists see advances in deep learning. *New York Times* **2012**.

122. Alvim-Gaston, M.; Grese, T.; Mahoui, A.; Palkowitz, A.; Pineiro-Nunez, M.; Watson, I. Open Innovation Drug Discovery (OIDD): A Potential Path to Novel Therapeutic Chemical Space. *Curr. Top. Med. Chem.* **2014**, *14*, 294–303. [CrossRef] [PubMed]

123. Margolin, A.A.; Bilal, E.; Huang, E.; Norman, T.C.; Ottestad, L.; Mecham, B.H.; Sauerwine, B.; Kellen, M.R.; Mangravite, L.M.; Furia, M.D.; et al. Systematic analysis of challenge-driven improvements in molecular prognostic models for breast cancer. *Sci. Transl. Med.* **2013**, *5*. [CrossRef] [PubMed]

124. Aksulu, A.; Wade, M. A Comprehensive Review and Synthesis of Open Source Research. *J. Assoc. Inf. Syst.* **2010**, *11*, 576–656. [CrossRef]

125. Stajich, J.E.; Lapp, H. Open source tools and toolkits for bioinformatics: Significance, and where are we? *Brief. Bioinform.* **2006**, *7*, 287–296. [CrossRef] [PubMed]

126. Kluyver, T.; Ragan-kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks—A publishing format for reproducible computational workflows. *Position. Power Acad. Publ. Play. Agents Agendas* **2016**, 87–90. [CrossRef]

127. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]

128. Demšar, J.; Curk, T.; Erjavec, A.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Staric, A.; Stajdohar, M.; et al. Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14*, 23492353.

129. Hall, M.A.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *SIGKDD Explor.* **2009**, *11*, 10–18. [CrossRef]

130. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), Savannah, Georgia, 22 August 2016; pp. 265–284.

131. Chollet, F. Keras. Available online: https://github.com/fchollet/keras (accessed on 23 April 2018).

132. Paszke, A.; Chanan, G.; Lin, Z.; Gross, S.; Yang, E.; Antiga, L.; Devito, Z. Automatic differentiation in PyTorch. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–4.

133. Yu, D.; Huang, X. Microsoft Computational Network Toolkit (CNTK). In *Neural Information Processing Systems (NIPS)*; Palais des Congrès de Montréal: Montréal, QC, Canada, 2015.

134. Vedaldi, A.; Lenc, K. MatConvNet—Convolutional Neural Networks for MATLAB. *arXiv* **2014**, arXiv:1412.4564.

135. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In Proceedings of the Third International AAAI Conference on Weblogs and Social Media, San Jose, CA, USA, 17–20 May 2009; pp. 361–362.

136. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]

137. Corsello, S.M.; Bittker, J.A.; Liu, Z.; Gould, J.; McCarren, P.; Hirschman, J.E.; Johnston, S.E.; Vrcic, A.; Wong, B.; Khan, M.; et al. The Drug Repurposing Hub: A next-generation drug library and information resource. *Nat. Med.* **2017**, *23*, 405–408. [CrossRef] [PubMed]

138. Jia, Z.; Liu, Y.; Guan, N.; Bo, X.; Luo, Z.; Barnes, M.R. 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*. [CrossRef] [PubMed]

139. Altae-Tran, H.; Ramsundar, B.; Pappu, A.S.; Pande, V. Low Data Drug Discovery with One-Shot Learning. *ACS Cent. Sci.* **2017**, *3*, 283–293. [CrossRef] [PubMed]

140. Zhou, H.; Gao, M.; Skolnick, J. Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci. Rep.* **2015**, *5*. [CrossRef] [PubMed]

141. Douguet, D. e-LEA3D: A computational-aided drug design web server. *Nucleic Acids Res.* **2010**, *38*. [CrossRef] [PubMed]

142. Miteva, M.A.; Guyon, F.; Tufféry, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **2010**, *38*. [CrossRef] [PubMed]

143. Zu, S.; Chen, T.; Li, S. Global optimization-based inference of chemogenomic features from drug-target interactions. *Bioinformatics* **2015**, *31*, 2523–2529. [CrossRef] [PubMed]

144. Louhimo, R.; Laakso, M.; Belitskin, D.; Klefström, J.; Lehtonen, R.; Hautaniemi, S. Data integration to prioritize drugs using genomics and curated data. *BioData Min.* **2016**, *9*. [CrossRef] [PubMed]

145. Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493–500. [CrossRef] [PubMed]

146. Duan, Q.; Reid, S.P.; Clark, N.R.; Wang, Z.; Fernandez, N.F.; Rouillard, A.D.; Readhead, B.; Tritsch, S.R.; Hodos, R.; Hafner, M.; et al. L1000CDS2: LINCS L1000 characteristic direction signatures search engine. *NPJ Syst. Biol. Appl.* **2016**, *2*, 16015. [CrossRef] [PubMed]

147. Carrella, D.; Napolitano, F.; Rispoli, R.; Miglietta, M.; Carissimo, A.; Cutillo, L.; Sirci, F.; Gregoretti, F.; Di Bernardo, D. Mantra 2.0: An online collaborative resource for drug mode of action and repurposing by network analysis. *Bioinformatics* **2014**, *30*, 1787–1788. [CrossRef] [PubMed]

148. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An Open chemical toolbox. *J. Cheminform.* **2011**, *3*. [CrossRef] [PubMed]

149. Williams, A.J.; Harland, L.; Groth, P.; Pettifer, S.; Chichester, C.; Willighagen, E.L.; Evelo, C.T.; Blomberg, N.; Ecker, G.; Goble, C.; et al. Open PHACTS: Semantic interoperability for drug discovery. *Drug Discov. Today* **2012**, *17*, 1188–1198. [CrossRef] [PubMed]

150. Garfield, E. The Who and Why of ISI. Essays of Information Scientist. 1966. Available online: http://garfield. library.upenn.edu/essays/V1p033y1962-73.pdf (accessed on 1 June 2018).

151. Repeat after me. *Nature* **2012**, *488*, 253. [CrossRef]

152. Begley, C.G.; Ellis, L.M. Drug development: Raise standards for preclinical cancer research. *Nature* **2012**, *483*, 531–533. [CrossRef] [PubMed]

153. Baker, M. Irreproducible biology research costs put at $28 billion per year. *Nature* **2015**. [CrossRef]

154. Peng, R.D. Reproducible research in computational science. *Science* **2011**, *334*, 1226–1227. [CrossRef] [PubMed]

155. Gentleman, R. Reproducible Research: A Bioinformatics Case Study. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*. [CrossRef] [PubMed]

156. Guney, E. Reproducible Drug Repurposing: When Similarity Does Not Suffice. *Pac. Symp. Biocomput.* **2016**, *22*, 132–143.

157. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]

158. Torre, D.; Krawczuk, P.; Jagodnik, K.M.; Lachmann, A.; Wang, Z.; Wang, L.; Kuleshov, M.V.; Ma'Ayan, A. Datasets2Tools, repository and search engine for bioinformatics datasets, tools and canned analyses. *Sci. Data* **2018**, *5*. [CrossRef] [PubMed]

159. Ram, K. Git can facilitate greater reproducibility and increased transparency in science. *Source Code Biol. Med.* **2013**, *8*. [CrossRef] [PubMed]

160. Merkel, D. Docker: Lightweight Linux containers for consistent development and deployment. *Linux J.* **2014**, *2014*, 2.