

Mining Complex Biomedical Literature for Actionable Knowledge on Rare Diseases



Vinicius M. Alves, Stephen J. Capuzzi, Nancy Baker, Eugene N. Muratov,
Alexander Trospsha, and Anthony J. Hickey

Introduction

Since the Enlightenment, the approach to physical, chemical, and biological processes has involved the development of analytical methods that would use quantitative data to build simple integrated models leading to prediction of drug action [1]. This was a logical approach as it could capture a large quantity of data in a manner that was easy to store and transmit from one individual to the another. However, as scientific research has become increasingly prominent type of human activity, there has been a dramatic growth of scientific publications reporting the research outcomes in a form combining both textual description and numerical data

V. M. Alves · S. J. Capuzzi

UNC Catalyst for Rare Diseases, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

N. Baker

ParlezChem, Hillsborough, NC, USA

E. N. Muratov

Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

Department of Pharmaceutical Sciences, Federal University of Paraiba, Joao Pessoa, PB, Brazil

A. Trospsha

Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

A. J. Hickey (✉)

UNC Catalyst for Rare Diseases, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

e-mail: ahickey@unc.edu

[2]. Most scientists do spend a lot of time thinking about the best verbal ways of describing and reasoning over their results and thus, a lot of useful information and knowledge could be obtained by reading the scientific literature. As important as reading is in the life of every scientist, the process of obtaining summative knowledge compiled from many publications is a non-scalable effort [3].

Fortunately, the advent of computer technology enables storing and efficient processing of large amounts of data, including textual sources. The analysis of this complex data allows mechanistic inferences to be drawn that promote novel hypotheses that illuminate fundamental natural phenomena. The importance of evolving computational methods that allow consideration of a wide range of data and their implications cannot be understated [4].

The current exponentially increasing cost and decades of inefficiency in drug discovery and development have resulted in a problematic situation with respect to pharmaceutical innovation and commercialization. The overwrought drug discovery pipeline may take up 15 years to develop a successful drug (considering hit-to-lead discovery/development, pre-clinical, and clinical studies), with an average cost estimated to be from \$800 million to \$1.5 billion [5]. This process is deemed as inadequate and unsustainable, especially as concerning its ability to provide a therapy for diseases that affect people in poor parts of the world, such as tropical diseases, as well as those affecting a limited number of patients, such as rare diseases, due to the potential resulting low revenue [6, 7]. A disruptive approach is required that can bring about revolutionary, not evolutionary, change equivalent to the changes that have occurred in the communications, electronics and financial industries over the last 25 years.

Rare diseases, which are defined as a condition that affects fewer than 200,000 people in the United States and 1 in 2000 people in the European Union, are particularly in need of disruptive and revolutionary drug discovery paradigms. Although, individually each rare disease affects a small portion of the total population only, their collective effect on the human population is substantial as there are over 7000 rare diseases that roughly affect 25–30 million people in the United States [8]. Alarmingly, very few patients can be treated with an approved medicine. Taken together, rare diseases represent a substantial burden on individuals, families, and whole economies [9, 10].

Developing a drug for a rare disease, on average, is half the cost of common diseases [11]. Still, considering the smaller amount of data and investment, any innovative approaches to treat these diseases will likewise be of value for drug discovery *writ large*. It is anticipated that once paradigms are developed for the rare diseases drug discovery, which have less financial benefit than more prevalent diseases, drug discovery efforts in general will become more efficient [12]. In addition to financial concerns and a limited patient populations, rare disease drug discovery also suffers from sparse and heterogeneous data [13], which hamper the ability to draw novel insights and treatment hypotheses. However, a growing number of rare diseases registries has been incorporate within in different databases [14, 15], such as Pharos [16] (<https://pharos.nih.gov/>), ClinVar [17, 18] (<https://www.ncbi.nlm.nih.gov/clinvar/>), the Online Mendelian Inheritance in Man (OMIM)

(<https://omim.org/>), among others [7]. While efforts have been made to promote the sharing of information between multi-disciplinary collaborations [19], there is still need to curate and properly integrate all of this information [13, 20–22].

Computational approaches have emerged as a practical solution to accelerate drug discovery efforts and reduce costs [23, 24]. One promising approach, named Literature-Based-Discovery (LBD), seeks to unlock biological observations hidden within informational sources, such as published texts and manuscripts [25]. Since 1988, when the relationship between magnesium and migraine was discovered in the literature by Swanson [26], other treatment hypotheses have been generated for many diseases, such as Parkinson's disease [27], multiple sclerosis [28], and cancer [29]. This approach has been also used to elucidate adverse drug effects [25, 30]. As such, LBD is a powerful new technology in the drug discovery arsenal.

In this chapter, we aim to review the status of available biomedical data on PubMed and describe how mining complex drug-target-disease relationships within this database could contribute to finding new targets, new repurposed medications, and novel drug candidates for rare diseases. The intent of the following discussion is to focus on the impact of consideration of complexity in drug discovery and clinical data to allow new therapies to emerge that can be rapidly screened and progressed to clinical application. The overall approach described may likely be one component of a strategy that will regenerate pharmaceutical development and promote a rational approach to the pharmacological element of health care delivery.

Biomedical Knowledge Data in the Scientific Literature

Bioactivity data such as the outcome of *in vivo* and *in vitro* assays have been growing extensively in publicly available repositories such as ChEMBL [31, 32] (<https://www.ebi.ac.uk/chembl/>) and PubChem [33, 34] (<http://pubchem.ncbi.nlm.nih.gov/>). Despite the growth of these databases, the scientific literature remains the largest repository of untapped biomedical data [2]. The United States National Library of Medicine (NLM) journal citation database (MEDLINE) is the preeminent source of biomedical literature, with ~30 million citations [35]. This database can be accessed through PubMed, a search engine maintained by NLM at the National Institutes of Health (NIH). It is possible to retrieve reference for scientific articles stored in MEDLINE by querying specific terms named Medical Subject Headings (MeSH) [36], which are used to index and categorize publications stored in MEDLINE. MeSH terms encompass most drugs, targets, and diseases present in scientific publications and could potentially be used to accelerate drug discovery [37].

The major approach to manipulate knowledge stored in the literature is through natural language processing, a subfield of artificial intelligence that allows computers to understand, interpret, and manipulate human language [38]. For this purpose, many dictionary-based systems that recognize passages in the literature with ontological terms have been proposed and evaluated [39]. The SciLite Annotations platform (<https://europepmc.org/Annotations>) provides means to link

research articles with biological data through text mining [40]. In a 2016 study, text mining on PubMed and social network analysis were integrated to analyze gene-gene interactions in order to identify new potential biomarkers for breast cancer [41]. More recently, text mining has been used to analyze gene-disease associations present in PubMed by integrating MeSH terms and co-occurrence methods [42].

Drug Repurposing

As discussed in the introduction, it may take a drug up to 15 years to reach the market [43]. This process usually includes discovery and development research, preclinical studies (*in vitro/in vivo* evaluation), and clinical research, divided in Phase I (safety and dose evaluation in healthy individuals), Phase II (efficacy and safety in small number of patients), Phase III (efficacy and safety in large number of patients), and Phase IV (post-market safety monitoring). During Phase II, approximately 90% of the compounds fail due to safety concerns and poor efficacy [44].

Drug repurposing, also known as repositioning or reprofiling, is a strategy to identify novel uses for approved or investigational drugs that are outside of the original therapeutic indication [45]. Recently, this approach has been a trending topic among researchers [46] and has attracted attention of companies due to the reduced cost associated with the low risk of failure, especially when safety evaluation has already been completed in preclinical and clinical trials [47]. Because repurposed drugs can skip safety evaluation during preclinical and Phase I studies, it is estimated that developing a repurposed drug costs on average only \$300 million over a 6.5 year period [48]. In addition to reduced cost and time, approximately 30% of repurposed drugs are approved, which can be seen as a market-oriented incentive to companies [45, 49]. For comparison, the typical approval rate for drugs entering clinical trials is 9.6% [50].

Repurposing studies very often are initiated after unexpected drug effects are observed during clinical trials or during pharmacovigilance upon their release on the market [51]. Many of the current repurposing studies have been initiated thanks to a serendipitous observation of unexpected drug effects upon clinical trials or following their release on the market. Prime examples of such discoveries are the stories of sildenafil (Viagra®) [52] and thalidomide [53, 54].

Recently, it has been shown in a bibliographic study [55] that more than 60% of all approved drugs or drug candidates (~35,000) have been tried in more than one disease, including 189 drugs that have been tried in >300 diseases each. Considering only approved drugs, more than 30% have been tested during their lifetime for at least one additional indication following their original approval [55]. Despite several success cases, drug repurposing still faces lack of financial support due to potentially low return, lower drug prices, and short patent duration [56, 57]. Nevertheless, this approach is still considered promising, especially for rare diseases [58]. Small grant programs to help develop drugs or treatments for rare diseases are usually available

from rare disease foundations [59]. The National Organization for Rare Disorders (NORD) (<http://rarediseases.org/>) provides recommendations to such organizations.

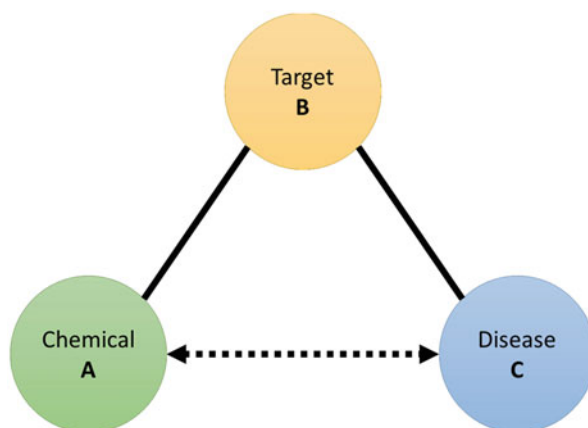
Using Chemotext to Infer Novel Therapies and Targets

Biological insights about the etiology of diseases, such as causative protein mutations or aberrant pathway signaling, and the potential drug treatments of these diseases are stored primarily in the biomedical literature [2]. As such, there exists biomedically relevant relationships between drugs, biological targets, and diseases, which we call DTD triangles, that lie latent within published texts [3, 60]. Using text-mining approaches, therefore, these DTD triangles can be identified and extracted from the published biomedical literature [61].

Text-mining capabilities in conjunction with the wealth of text-based data stored within PubMed considerations led to the development of Chemotext [62], a computational algorithm which extracts MeSH terms describing “drugs”, “targets”, and “diseases” and generates DTD triangles. Chemotext is based on the frequency with which MeSH terms of interest co-occur in abstracts of papers annotated in PubMed. Chemotext is thus an extension of Swanson’s ABC paradigm wherein “A” terms are drug (chemical) MeSH terms, “B” terms are target-associated MeSH terms, i.e., proteins and pathways, and “C” terms are disease MeSH terms (Fig. 1).

The underlying DTD triangle generation starts with the observation that the MeSH term of drug “A” co-occurs in the same articles as the MeSH term of target “B” while the MeSH term of disease “C” co-occurs in the same or additional articles with the same target B. Thus, if drug A and disease C have not been mentioned together in the same article, an “A–C” connection mediated through target B can be inferred, completing a DTD triangle. This analysis, enabled by the Chemotext approach, leads to the identification of a new possible therapeutic use of drug “A”.

Fig. 1 Swanson’s ABC paradigm incorporated in Chemotext. Chemical A is proposed to affect disease C since both terms are associated with target B. Solid lines (edges) indicate an actual text-based relationship, while dashed lines (edges) indicate proposed connections



The power and efficacy of Chemotext is demonstrated by elucidation of the antineoplastic agent imatinib as a potential drug repurposing candidate for the treatment of severe refractory asthma. Imatinib is an FDA-approved tyrosine kinase inhibitor that is used in the treatment chronic myeloid leukemia (CML). Imatinib inhibits the activity of KIT, which reduces bone marrow mast-cell numbers in patients with CML [63]. KIT is also present in lung mast cells and was hypothesized as a basis of the pathobiology of severe refractory asthma [64], which is characterized by an adverse response to traditional glucocorticoid asthma treatment [65]. Figure 2 shows how Chemotext can be used to link Imatinib (A), Proto-Oncogene Proteins c-kit (B), and asthma (C).

In 2017, a proof-of-principle trial demonstrated that imatinib reduced airway hyperresponsiveness, a physiological marker of severe asthma, as well as on airway mast-cell numbers and activation in patients with severe asthma. Since this publication had not yet been entered into the.

MEDLINE database, it was used a validation test of the Chemotext algorithm. Through co-occurrences of these MeSH terms in previously published studies, Chemotext was used to draw the interference between imatinib, KIT, and asthma, which constitutes a DTD triangle (Fig. 2). This case study demonstrates that Chemotext can identify drug repurposing candidates and targets through text-based inferences alone.

Mining Other Sources of Biomedical Data for Drug Repurposing

Mining literature data can afford rapid identification of all published studies that could confirm connections between drugs, their targets, underlying biological pathways, and diseases, including enabling new inferences of such connections [3, 60]. The elucidation of the mechanistic relationships between these connections is at the core of modern drug discovery research [61]. Currently, there are several databases with valuable information for drug discovery that could be connected to complete a DTD triangle. ChEMBL [31, 32] (<https://www.ebi.ac.uk/chembl/>) and PubChem [33, 34] (<http://pubchem.ncbi.nlm.nih.gov/>) contain many chemical–target (“A–B”) and chemical–disease (“A–C”) relationships. Other databases contain target–disease (“B–C”) associations, such as ClinVar [17, 18] (<https://www.ncbi.nlm.nih.gov/clinvar/>), the Online Mendelian Inheritance in Man (OMIM) (<https://omim.org/>). Pharos [16] (<https://pharos.nih.gov/>), specifically, contains data on the whole DTD triangle for many diseases. Several databases are available containing parts of the triangle available for rare diseases, such as Malacards [66] (<http://www.malacards.org/>) the National Organization for Rare Disorders (NORD) [67] (<https://rarediseases.org/>), the Genetic and Rare Diseases Information Center (GARD) [68]

Chemotext

Home Find Connected Terms **Find Shared Terms** Path Search Find Articles

Imatinib

Asthma

Search

Click to Include Subterms

Terms	Same Article	Imatinib Only	Asthma Only
Allergens	3	1	7806
Stem Cell Factor	3	73	28
Ovalbumin	2	3	2258
Proto-Oncogene Proteins c-kit	2	955	14
Glucocorticoids	1	26	3141
Cytokines	1	56	2804
Biomarkers	1	65	1896
Protein-Tyrosine Kinases	1	1509	74
Immunoglobulin G	1	12	1335
Interleukin-13	1	1	830

10 Set Table Limit →

Total Count: 1281

View CSV CSV with pmids

Proteins-Pathways-Intermediaries-Other ▼

Fig. 2 Example showing how Chemotext connects Imatinib and Asthma with shared terms. In this example, query terms “Imatinib” and “Asthma” were searched in the Find Shared Terms module. The list of full associations was filtered by Proteins-Pathways-Intermediaries-Other. The MeSH term “Proto- Oncogene Proteins c-kit” was the fourth highest ranked shared term (two shared articles) selected as the potential biological target in the clinical outcome pathway

(<https://rarediseases.info.nih.gov/>), and the Infohub for Rare Diseases (<https://rarediseases.oscar.ncsu.edu/>).

Recently, NIH has launched the Biomedical Data Translator program (<https://ncats.nih.gov/translator>), which has integrated many data sources with multiple

types of content, such as diseases, patient-reported outcomes, electronic health records, microbiome, proteins, genes, chemicals, among others. This massive project attempts to integrate currently available medical research data towards accelerated development of new treatments. The major challenge to establish valuable connections, as in any data science project, is proper curation of the data [13, 20–22]. To establish useful relationships between these sources of data, knowledge graphs have emerged as a practical solution. A knowledge graph is a network of entities that acquires and integrates information into an ontology and applies a reasoner to derive new knowledge [68]. A 2016 study has applied network-based modeling within to identify promising multi-target drugs for triple negative breast cancer [11]. More recently, a study has applied knowledge graphs to integrate different data sources on diseases and drugs to suggest the repurposing of 21 drugs for Autosomal Dominant Polycystic Kidney Disease (ADPKD) [68].

There has also been a growing interest in using social media to supplement established approaches for pharmacovigilance [69, 70]. The use of social media, also called “social listening”, therefore, is a potential resource for repurposing. Social media has been recently used in public health to estimate trends of cholera outbreak in the aftermath of the 2010 earthquake in Haiti [71], seasonal influenza surveillance [72], and onset of mental illness [73]. As previously discussed, many repurposed drugs have been discovered through adverse side effects observed during clinical trials or pharmacovigilance. Many people have used social media to report adverse effects of their medications. Several studies analyzing adverse reactions on social media have been published recently [30, 74, 75], which makes social media a potential source of adverse effect data to be mined for repurposing.

Drug Repurposing and Bibliometric Analysis on Rare Diseases

Several repurposing stories for rare diseases have been reported in the recent years. For instance, metformin has been studied to treat idiopathic pulmonary fibrosis [76]. A recent study suggests that inhibitors of p110 β , a catalytic subunit of the phosphoinositide 3-kinase (PI3K) gene family, commonly associated with cancer, might prevent cognitive and behavioral defects and become a promising disease-modifying strategy for fragile X syndrome and other brain disorders [77]. Fenfluramine, initially proposed as an appetite suppressant and withdrawn from the market, has been submitted to the FDA for the treatment of Dravet syndrome [78].

Many computational approaches historically applied for drug discovery, such as quantitative structure-activity relationships (QSAR) modeling, similarity search, molecular docking, etc., have been successfully applied for drug repurposing as well [79]. Computational drug repurposing approaches have been widely applied to

neglected tropical diseases [80–84], and, more recently, to rare diseases [58, 83]. The eMatchSite, a platform for compare drug-binding sites have been applied to propose the possibility to repurpose a steroidal aromatase inhibitor to treat Niemann-Pick disease type C [85]. A structure-based virtual screening approach has been applied to screen FDA approved drugs on ENGase, a potential target for the treatment of N-Glycanase (NGLY1) deficiency. The authors experimentally confirmed the activity of rabepazole (IC₅₀ = 4.4. μ M) on ENGase as a promising treatment to patients suffering from NGLY1 deficiency [69].

Mining literature data allows the exploitation of opportunities to reposition known drugs interacting with proteins associated with diseases [3, 60]. The integration of data on drug-target-disease to form networks has become a valuable approach for computational drug repositioning research [86]. Recently, a study has used bioinformatics methods and bibliographic research to propose the repositioning of some drugs as potential competitors against idiopathic pulmonary fibrosis [87].

As of June of 2019, there are 244,911 references with the term “rare disease” through the text and 17,134 references with the term “rare disease” in the title or abstract. Here, we performed a brief bibliometric analysis on drug repurposing for rare diseases, similar to the one that was recently published by Baker et al. [55]. We mined PubMed using earlier text-mining work [37] to identify articles in PubMed where a chemical entity was described in terms of its therapeutic association with a rare disease. We determined this relationship by examining the MeSH annotations in a stepwise manner (described in the supplementary material online). All drug-disease combinations were extracted, along with the year the article was published, into a separate dataset. This set included citations with no abstract and those in languages other than English, as long as they were annotated, and the annotations met the criteria.

In our analysis, we found that only 1267 out of more than 7000 rare diseases have been studied in association with a chemical entity. It was known that only a small fraction of rare diseases has associated treatments, but our findings reveal there is still a major gap in research for rare diseases, since many of them have not been associated with any chemical entity as a potential treatment. These findings reinforce the need to expand research on the development of novel therapies for rare diseases. As one can see in Fig. 3, 6570 out of 12,376 chemicals (53%) have been associated with only one rare disease, while 4796 (38%) have been associated with two to ten diseases, 984 (7.0%) have been associated with eleven to 100 diseases, and 26 (0.20%) chemicals with more than 100 diseases.

We show in Table 1 the top 30 drugs that were tested for rare diseases. Sixteen out of 30 were among the top drugs most tested in the previous study [55]. As one can see, most of these drugs are used to suppress the immune system and/or to decrease inflammation, such as glucocorticoid medications (prednisone, prednisolone, dexamethasone, methylprednisolone, hydrocortisone, and cortisone), cancer chemotherapy agents (cyclophosphamide, bevacizumab, methotrexate), and medications used to prevent transplant rejections (sirolimus, rituximab, cyclosporine). The rare

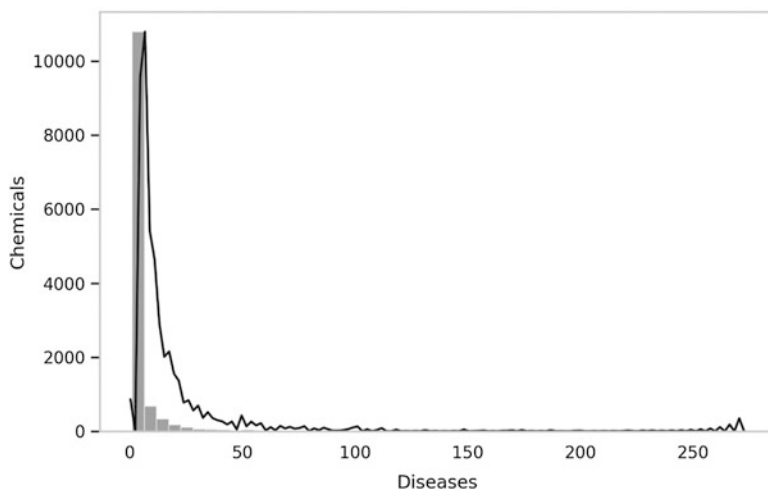


Fig. 3 Distribution of chemicals tested in rare diseases mined from PubMed

diseases with most publications and chemicals tested are presented in Table 2. Most of these diseases are rare forms of cancer, such as sarcoma, and neoplasm, and multiple forms of carcinoma, which explains why most of the most studied drugs present in Table 1 are suppressant of immune system, anti-inflammatory, and anti-cancer drugs. Surprisingly, none of the most studied drugs were used in some of the most studied diseases, such as malaria, tuberculosis, and Alzheimer.

Final Remarks

There is an urgent need for the development of treatments or cures for rare diseases. The complex biological systems and nature of drug discovery make iterative mechanistic strategies costly and inefficient. Current developments in database development, text mining, and machine learning tools allows efficient and inexpensive navigation through inferences to the identification of novel or repurposed drug candidates. The same principles can be employed to the traverse the complexity of drug delivery systems and biopharmaceutical principles that result in optimal drug disposition to achieve the desired therapeutic effect. In this manner, the development

Table 1 Top 30 drugs most tested in rare diseases with publications count

Chemicals	Rare diseases count	Publications count
Prednisolone	272	1627
Prednisone	233	1857
Dexamethasone	229	1598
Methylprednisolone	221	1162
Cyclophosphamide	199	2309
Cyclosporine	187	1376
Rituximab	174	2123
Methotrexate	170	1790
Interferon-alpha	167	3316
Immunoglobulin G	156	782
Sirolimus	149	802
Ascorbic Acid	148	504
Vitamin E	131	635
Infliximab	119	2467
Adrenocorticotrophic Hormone	118	969
Tretinoin	112	776
Tacrolimus	112	416
Thalidomide	111	1320
Hydrocortisone	111	345
Aspirin	110	693
Heparin	106	396
Indomethacin	105	941
Curcumin	102	590
Bevacizumab	101	1271
Granulocyte Colony-Stimulating Factor	101	1253
Interferon-gamma	101	723
Cortisone	100	723
Acetylcysteine	100	315
Pentoxifylline	100	262

of novel pharmaceutical treatment options can focus on the generation of data suited to regulatory scrutiny and positive clinical outcomes without investment in the tangential iterative data generation that has historically been required to support statistical validation of the action, process, or clinical observations that surround the optimal approach.

Table 2 Top 30 rare diseases ranked by number of publications and chemicals tested

Rare disease	Publications	Chemicals
Adenocarcinoma	11811	2226
Liver Neoplasms	10773	2063
Carcinoma, Non-Small-Cell Lung	7527	1355
Carcinoma, Hepatocellular	6570	1647
Alzheimer Disease	5864	1500
Multiple Myeloma	5680	873
Malaria	5620	667
Urinary Bladder Neoplasms	5486	824
Leukemia, Myelogenous, Chronic, BCR-ABL Positive	5173	504
Leukemia, Myeloid, Acute	5066	897
Crohn Disease	4551	342
Kidney Neoplasms	4148	841
Tuberculosis	3876	535
Hemophilia A	3732	184
Carcinoma, Renal Cell	3193	627
Hodgkin Disease	2212	320
Cystic Fibrosis	2070	346
Cytomegalovirus Infections	2058	220
Neuroblastoma	2051	770
Osteosarcoma	2012	671
Myelodysplastic Syndromes	2010	235
Encephalomyelitis, Autoimmune, Experimental	1911	858
Anemia, Aplastic	1714	195
Sarcoma	1696	433
Spondylitis, Ankylosing	1667	171
Leprosy	1606	176
Fatty Liver	1603	651
Respiratory Distress Syndrome, Newborn	1530	180
Pulmonary Fibrosis	1459	574
Leishmaniasis, Visceral	1453	258

References

1. Hickey, A.J., and H.D.C. Smyth. 2011. *Pharmaco-complexity*. Boston: Springer US.
2. Hunter, L.E. 2017. Knowledge-based biomedical data science. *Data Science Journal* 1: 1–7. <https://doi.org/10.3233/DS-170001>.
3. Przybyła, P., M. Shardlow, S. Aubin, R. Bossy, R. Eckart de Castilho, S. Piperidis, J. McNaught, and S. Ananiadou. 2016. Text mining resources for the life sciences. *Database* 2016. <https://doi.org/10.1093/database/baw145>.
4. Pan, W., Z. Li, Y. Zhang, and C. Weng. 2018. The new hardware development trend and the challenges in data management and analysis. *Data Science and Engineering* 3: 263–276. <https://doi.org/10.1007/s41019-018-0072-6>.

5. DiMasi, J.A., H.G. Grabowski, and R.W. Hansen. 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics* 47: 20–33. <https://doi.org/10.1016/J.JHEALECO.2016.01.012>.
6. Baxter, K., E. Horn, N. Gal-Edd, K. Zonno, J. O’Leary, P.F. Terry, and S.F. Terry. 2013. An end to the myth: There is no drug development pipeline. *Science Translational Medicine* 5: 171cm1. <https://doi.org/10.1126/scitranslmed.3003505>.
7. Zhao, M., and D.-Q.Q. Wei. 2018. Rare diseases: Drug discovery and informatics resource. *Interdisciplinary Sciences: Computational Life Sciences* 10: 195–204. <https://doi.org/10.1007/s12539-017-0270-3>.
8. Valdez, R., L. Ouyang, and J. Bolen. 2016. Public health and rare diseases: oxymoron no more. *Preventing Chronic Disease* 13: 150491. <https://doi.org/10.5888/pcd13.150491>.
9. Kakkis, E.D., M. O’Donovan, G. Cox, M. Hayes, F. Goodsaid, P. Tandon, P. Furlong, S. Boynton, M. Bozic, M. Orfali, and M. Thornton. 2015. Recommendations for the development of rare disease drugs using the accelerated approval pathway and for qualifying biomarkers as primary endpoints. *Orphanet Journal of Rare Diseases* 10: 16. <https://doi.org/10.1186/s13023-014-0195-4>.
10. Angelis, A., D. Tordrup, and P. Kanavos. 2015. Socio-economic burden of rare diseases: A systematic review of cost of illness evidence. *Health Policy* 119: 964–979. <https://doi.org/10.1016/j.healthpol.2014.12.016>.
11. Vitali, F., L.D. Cohen, A. Demartini, A. Amato, V. Eterno, A. Zambelli, and R. Bellazzi. 2016. A network- based data integration approach to support drug repurposing and multi-target therapies in triple negative breast cancer. *PLoS One* 11: e0162407. <https://doi.org/10.1371/journal.pone.0162407>.
12. Ekins, S. 2017. Industrializing rare disease therapy discovery and development. *Nature Biotechnology* 35: 117–118. <https://doi.org/10.1038/nbt.3787>.
13. Roos M, López Martin E, Wilkinson MD (2017) Preparing data at the source to Foster interoperability across rare disease resources. In: de la Posada Paz M, Taruscio D, Groft S (eds) Rare diseases epidemiology: Update and overview. Advances in Experimental Medicine and Biology. Springer, Cham, pp 165–179
14. Kodra, Y., M. Posada de la Paz, A. Coi, M. Santoro, F. Bianchi, F. Ahmed, Y.R. Rubinstein, J. Weinbach, and D. Taruscio. 2017. Data quality in rare diseases registries. In *Advances in experimental medicine and biology*, 149–164. Cham: Springer.
15. Litterman, N.K., M. Rhee, D.C. Swinney, and S. Ekins. 2014. Collaboration for rare disease drug discovery research. *F1000Research* 3: 261. <https://doi.org/10.12688/f1000research.5564.1>.
16. Nguyen, D.T., S. Mathias, C. Bologna, S. Brunak, N. Fernandez, A. Gaulton, A. Hersey, J. Holmes, L.J. Jensen, A. Karlsson, G. Liu, A. Ma’ayan, G. Mandava, S. Mani, S. Mehta, J. Overington, J. Patel, A.D. Rouillard, S. Schürer, T. Sheils, A. Simeonov, L.A. Sklar, N. Southall, O. Ursu, D. Vidovic, A. Waller, J. Yang, A. Jadhav, T.I. Oprea, and R. Guha. 2017. Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Research* 45: D995–D1002. <https://doi.org/10.1093/nar/gkw1072>.
17. Landrum, M.J., J.M. Lee, M. Benson, G.R. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, K. Karapetyan, K. Katz, C. Liu, Z. Maddipatla, A. Malheiro, K. McDaniel, M. Ovetsky, G. Riley, G. Zhou, J.B. Holmes, B.L. Kattman, and D.R. Maglott. 2018. ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research* 46: D1062–D1067. <https://doi.org/10.1093/nar/gkx1153>.
18. Landrum, M.J., J.M. Lee, G.R. Riley, W. Jang, W.S. Rubinstein, D.M. Church, and D.R. Maglott. 2014. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research* 42: D980–D985. <https://doi.org/10.1093/nar/gkt1113>.
19. Kaufmann, P., A.R. Pariser, and C. Austin. 2018. From scientific discovery to treatments for rare diseases – The view from the National Center for Advancing Translational Sciences –

- Office of Rare Diseases Research. *Orphanet Journal of Rare Diseases* 13: 196. <https://doi.org/10.1186/s13023-018-0936-x>.
20. Fourches, D., E. Muratov, and A. Tropsha. 2010. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling* 50: 1189–1204. <https://doi.org/10.1021/ci100176x>.
 21. ———. 2016. Trust, but Verify II: A practical guide to chemogenomics data curation. *Journal of Chemical Information and Modeling* 56: 1243–1252. <https://doi.org/10.1021/acs.jcim.6b00129>.
 22. ———. 2015. Curation of chemogenomics data. *Nature Chemical Biology* 11: 535–535. <https://doi.org/10.1038/nchembio.1881>.
 23. Rognan, D. 2017. The impact of in silico screening in the discovery of novel and safer drug candidates. *Pharmacology & Therapeutics* 175: 47–66. <https://doi.org/10.1016/j.pharmthera.2017.02.034>.
 24. Makhouri, F.R., and J.B. Ghasemi. 2019. Combating diseases with computational strategies used for drug design and discovery. *Current Topics in Medicinal Chemistry* 18: 2743–2773. <https://doi.org/10.2174/1568026619666190121125106>.
 25. Henry, S., and B.T. McInnes. 2017. Literature based discovery: Models, methods, and trends. *Journal of Biomedical Informatics* 74: 20–32. <https://doi.org/10.1016/j.jbi.2017.08.011>.
 26. Swanson, D.R. 1988. Migraine and magnesium: Eleven neglected connections. *Perspectives in Biology and Medicine* 31: 526–557.
 27. Kostoff, R.N., and M.B. Briggs. 2008. Literature-Related Discovery (LRD): Potential treatments for Parkinson’s disease. *Technological Forecasting and Social Change* 75: 226–238. <https://doi.org/10.1016/j.techfore.2007.11.007>.
 28. Kostoff, R.N., M.B. Briggs, and T.J. Lyons. 2008. Literature-related discovery (LRD): Potential treatments for multiple sclerosis. *Technological Forecasting and Social Change* 75: 239–255. <https://doi.org/10.1016/j.techfore.2007.11.002>.
 29. Choi, B.-K., T. Dayaram, N. Parikh, A.D. Wilkins, M. Nagarajan, I.B. Novikov, B.J. Bachman, S.Y. Jung, P.J. Haas, J.L. Labrie, C.R. Pickering, A.K. Adikesavan, S. Regenbogen, L. Kato, A. Lelescu, C.M. Buchovecky, H. Zhang, S.H. Bao, S. Boyer, G. Weber, K.L. Scott, Y. Chen, S. Spangler, L.A. Donehower, and O. Lichtarge. 2018. Literature-based automated discovery of tumor suppressor p53 phosphorylation and inhibition by NEK2. *Proceedings of the National Academy of Sciences* 115: 10666–10671. <https://doi.org/10.1073/pnas.1806643115>.
 30. La, M.K., A. Sedykh, D. Fourches, E. Muratov, and A. Tropsha. 2018. Predicting adverse drug effects from literature- and database-mined assertions. *Drug Safety* 41: 1059–1072. <https://doi.org/10.1007/s40264-018-0688-5>.
 31. Willighagen, E.L., A. Waagmeester, O. Spjuth, P. Ansell, A.J. Williams, V. Tkachenko, J. Hastings, B. Chen, and D.J. Wild. 2013. The ChEMBL database as linked open data. *Journal of Cheminformatics* 5: 23. <https://doi.org/10.1186/1758-2946-5-23>.
 32. Gaulton, A., A. Hersey, M. Nowotka, A.P. Bento, J. Chambers, D. Mendez, P. Motow, F. Atkinson, L.J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M.P. Magariños, J.P. Overington, G. Papadatos, I. Smit, and A.R. Leach. 2017. The ChEMBL database in 2017. *Nucleic Acids Research* 45: D945–D954. <https://doi.org/10.1093/nar/gkw1074>.
 33. Wang, Y., T. Suzek, J. Zhang, J. Wang, S. He, T. Cheng, B.A. Shoemaker, A. Gindulyte, and S.H. Bryant. 2014. PubChem BioAssay: 2014 update. *Nucleic Acids Research* 42: D1075–D1082. <https://doi.org/10.1093/nar/gkt978>.
 34. Wang, Y., J. Xiao, T.O. Suzek, J. Zhang, J. Wang, Z. Zhou, L. Han, K. Karapetyan, S. Dracheva, B.A. Shoemaker, E. Bolton, A. Gindulyte, and S.H. Bryant. 2012. PubChem’s BioAssay database. *Nucleic Acids Research* 40: D400–D412. <https://doi.org/10.1093/nar/gkr1132>.
 35. Roberts, R.J. 2001. PubMed central: The GenBank of the published literature. *Proceedings of the National Academy of Sciences* 98: 381–382. <https://doi.org/10.1073/pnas.98.2.381>.

36. NLM. 2019. *Medical subject headings*. <https://www.nlm.nih.gov/mesh/meshhome.html>. Accessed 3 Jun 2019.
37. Baker, N.C., and B.M. Hemminger. 2010. Mining connections between chemicals, proteins, and diseases extracted from Medline annotations. *Journal of Biomedical Informatics* 43: 510–519. <https://doi.org/10.1016/j.jbi.2010.03.008>.
38. Kreimeyer, K., M. Foster, A. Pandey, N. Arya, G. Halford, S.F. Jones, R. Forshee, M. Walderhaug, and T. Botsis. 2017. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. *Journal of Biomedical Informatics* 73: 14–29. <https://doi.org/10.1016/j.jbi.2017.07.012>.
39. Funk, C., W. Baumgartner, B. Garcia, C. Roeder, M. Bada, K.B. Cohen, L.E. Hunter, and K. Verspoor. 2014. Large-scale biomedical concept recognition: An evaluation of current automatic annotators and their parameters. *BMC Bioinformatics* 15: 59. <https://doi.org/10.1186/1471-2105-15-59>.
40. Venkatesan, A., J.-H. Kim, F. Talo, M. Ide-Smith, J. Gobeill, J. Carter, R. Batista-Navarro, S. Ananiadou, P. Ruch, and J. McEntyre. 2016. SciLite: A platform for displaying text-mined annotations as a means to link research articles with biological data. *Wellcome Open Research* 1: 25. <https://doi.org/10.12688/wellcomeopenres.10210.2>.
41. Jurca, G., O. Addam, A. Aksac, S. Gao, T. Özyer, D. Demetrick, and R. Alhaji. 2016. Integrating text mining, data mining, and network analysis for identifying genetic breast cancer trends. *BMC Research Notes* 9: 236. <https://doi.org/10.1186/s13104-016-2023-5>.
42. Zhou, J., and B.-Q. Fu. 2018. The research on gene-disease association based on text-mining of PubMed. *BMC Bioinformatics* 19: 37. <https://doi.org/10.1186/s12859-018-2048-y>.
43. IFPMA. 2017. *The pharmaceutical industry and global health: Facts and figures*. <https://www.ifpma.org/wp-content/uploads/2017/02/IFPMA-Facts-And-Figures-2017.pdf>. Accessed 7 Jun 2019.
44. Arrowsmith, J. 2011. Trial watch: Phase II failures: 2008–2010. *Nature Reviews. Drug Discovery* 10: 328–329. <https://doi.org/10.1038/nrd3439>.
45. Ashburn, T.T., and K.B. Thor. 2004. Drug repositioning: Identifying and developing new uses for existing drugs. *Nature Reviews. Drug Discovery* 3: 673–683. <https://doi.org/10.1038/nrd1468>.
46. Langedijk, J., A.K. Mantel-Teeuwisse, D.S. Slijkerman, and M.-H.D.B. Schutjens. 2015. Drug repositioning and repurposing: Terminology and definitions in literature. *Drug Discovery Today* 20: 1027–1034. <https://doi.org/10.1016/j.drudis.2015.05.001>.
47. Cha, Y., T. Erez, I.J. Reynolds, D. Kumar, J. Ross, G. Koytiger, R. Kusko, B. Zeskind, S. Risso, E. Kagan, S. Papapetropoulos, I. Grossman, and D. Laifenfeld. 2018. Drug repurposing from the perspective of pharmaceutical companies. *British Journal of Pharmacology* 175: 168–180. <https://doi.org/10.1111/bph.13798>.
48. Nosengo, N. 2016. Can you teach old drugs new tricks? *Nature* 534: 314–316. <https://doi.org/10.1038/534314a>.
49. Hernandez, J.J., M. Pryszyk, L. Smith, C. Yanchus, N. Kurji, V.M. Shahani, and S.V. Molinski. 2017. Giving drugs a second chance: Overcoming regulatory and financial hurdles in repurposing approved drugs as cancer therapeutics. *Frontiers in Oncology* 7: 273. <https://doi.org/10.3389/fonc.2017.00273>.
50. Bio. 2016. *Clinical development success rates*. https://www.bio.org/sites/default/files/Clinical_Development_Success_Rates_2006-2015_-_BIO,_Biomedtracker,_Amplion_2016.pdf. Accessed 19 Jun 2019.
51. Pushpakom, S., F. Iorio, P.A. Eyers, K.J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, and M. Pirmohamed. 2019. Drug repurposing: Progress, challenges and recommendations. *Nature Reviews. Drug Discovery* 18: 41–58. <https://doi.org/10.1038/nrd.2018.168>.
52. Langtry, H.D., and A. Markham. 1999. Sildenafil. *Drugs* 57: 967–989. <https://doi.org/10.2165/00003495-199957060-00015>.

53. Franks, M.E., G.R. Macpherson, and W.D. Figg. 2004. Thalidomide. *Lancet* 363: 1802–1811. [https://doi.org/10.1016/S0140-6736\(04\)16308-3](https://doi.org/10.1016/S0140-6736(04)16308-3).
54. NCI. 2006. *Thalidomide*. <https://www.cancer.gov/about-cancer/treatment/drugs/thalidomide?redirect=true>. Accessed 7 Jun 2019.
55. Baker, N.C., S. Ekins, A.J. Williams, and A. Tropsha. 2018. A bibliometric review of drug repurposing. *Drug Discovery Today* 23: 661–672. <https://doi.org/10.1016/j.drudis.2018.01.018>.
56. Novac, N. 2013. Challenges and opportunities of drug repositioning. *Trends in Pharmacological Sciences* 34: 267–272. <https://doi.org/10.1016/j.tips.2013.03.004>.
57. Ding, X. 2016. Drug repositioning needs a rethink. *Nature* 535: 355–355. <https://doi.org/10.1038/535355d>.
58. Delavan, B., R. Roberts, R. Huang, W. Bao, W. Tong, and Z. Liu. 2018. Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discovery Today* 23: 382–394. <https://doi.org/10.1016/j.drudis.2017.10.009>.
59. Sun, W., W. Zheng, and A. Simeonov. 2017. Drug discovery and development for rare genetic disorders. *American Journal of Medical Genetics. Part A* 173: 2307–2322. <https://doi.org/10.1002/ajmg.a.38326>.
60. Wei, C.-H., H.-Y. Kao, and Z. Lu. 2013. PubTator: A web-based text mining tool for assisting biocuration. *Nucleic Acids Research* 41: W518–W522. <https://doi.org/10.1093/nar/gkt441>.
61. Hughes, J.P., S. Rees, S.B. Kalindjian, and K.L. Philpott. 2011. Principles of early drug discovery. *British Journal of Pharmacology* 162: 1239–1249. <https://doi.org/10.1111/j.1476-5381.2010.01127.x>.
62. Capuzzi, S.J., T.E. Thornton, K. Liu, N. Baker, W.I. Lam, C. O'Banion, E.N. Muratov, D. Pozefsky, A. Tropsha, C.P. O'Banion, E.N. Muratov, D. Pozefsky, and A. Tropsha. 2018. Chemotext: A publicly available web server for mining drug–target–disease relationships in PubMed. *Journal of Chemical Information and Modeling* 58: 212–218. <https://doi.org/10.1021/acs.jcim.7b00589>.
63. Reichardt, P. 2018. The story of Imatinib in GIST – A journey through the development of a targeted therapy. *Oncology Research Treatment* 41: 472–477. <https://doi.org/10.1159/000487511>.
64. Fuehrer, N.E., A.M. Marchevsky, and J. Jagirdar. 2009. Presence of c-KIT-positive mast cells in obliterative bronchiolitis from diverse causes. *Archives of Pathology & Laboratory Medicine* 133: 1420–1425. <https://doi.org/10.1043/1543-2165-133.9.1420>.
65. Cahill, K.N., H.R. Katz, J. Cui, J. Lai, S. Kazani, A. Crosby-Thompson, D. Garofalo, M. Castro, N. Jarjour, E. DiMango, S. Erzurum, J.L. Trevor, K. Shenoy, V.M. Chinchilli, M.E. Wechsler, T.M. Laidlaw, J.A. Boyce, and E. Israel. 2017. KIT inhibition by Imatinib in patients with severe refractory asthma. *The New England Journal of Medicine* 376: 1911–1920. <https://doi.org/10.1056/NEJMoa1613125>.
66. Rappaport, N., M. Twik, I. Plaschkes, R. Nudel, T.I. Stein, J. Levitt, M. Gershoni, C.P. Morrey, M. Safran, and D. Lancet. 2017. MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Research* 45: D877–D887. <https://doi.org/10.1093/nar/gkw1012>.
67. Putkowski, S. 2010. The National Organization for Rare Disorders (NORD). *NASN School Nurse* 25: 38–41. <https://doi.org/10.1177/1942602X09352796>.
68. Lewis, J., M. Snyder, and H. Hyatt-Knorr. 2017. Marking 15 years of the genetic and rare diseases information center. *Translational Science of Rare Diseases* 2: 77–88. <https://doi.org/10.3233/TRD-170011>.
69. Bi, Y., M. Might, H. Vankayalapati, and B. Kuberan. 2017. Repurposing of proton pump inhibitors as first identified small molecule inhibitors of endo- β -N-acetylglucosaminidase (ENGase) for the treatment of NGLY1 deficiency, a rare genetic disease. *Bioorganic & Medicinal Chemistry Letters* 27: 2962–2966. <https://doi.org/10.1016/j.bmcl.2017.05.010>.
70. Tricco, A.C., W. Zarin, E. Lillie, S. Jeblee, R. Warren, P.A. Khan, R. Robson, B. Pham, G. Hirst, and S.E. Straus. 2018. Utility of social media and crowd-intelligence data for

- pharmacovigilance: A scoping review. *BMC Medical Informatics and Decision Making* 18: 38. <https://doi.org/10.1186/s12911-018-0621-y>.
71. Chunara, R., J.R. Andrews, and J.S. Brownstein. 2012. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American Journal of Tropical Medicine and Hygiene* 86: 39–45. <https://doi.org/10.4269/ajtmh.2012.11-0597>.
72. Kagashe, I., Z. Yan, and I. Suheryani. 2017. Enhancing seasonal influenza surveillance: Topic analysis of widely used medicinal drugs using twitter data. *Journal of Medical Internet Research* 19: e315. <https://doi.org/10.2196/jmir.7393>.
73. Reece, A.G., A.J. Reagan, K.L.M. Lix, P.S. Dodds, C.M. Danforth, and E.J. Langer. 2017. Forecasting the onset and course of mental illness with twitter data. *Scientific Reports* 7: 13006. <https://doi.org/10.1038/s41598-017-12961-9>.
74. Adrover, C., T. Bodnar, Z. Huang, A. Telenti, and M. Salathé. 2015. Identifying adverse effects of HIV drug treatment and associated sentiments using twitter. *JMIR Public Health and Surveillance* 1: e7. <https://doi.org/10.2196/publichealth.4488>.
75. MacKinlay, A., H. Aamer, and A.J. Yepes. 2017. Detection of adverse drug reactions using medical named entities on twitter. *AMIA Annual Symposium Proceedings. AMIA Symposium 2017*: 1215–1224.
76. Rangarajan, S., N.B. Bone, A.A. Zmijewska, S. Jiang, D.W. Park, K. Bernard, M.L. Locy, S. Ravi, J. Deshane, R.B. Mannon, E. Abraham, V. Darley-USmar, V.J. Thannickal, and J.W. Zmijewski. 2018. Metformin reverses established lung fibrosis in a bleomycin model. *Nature Medicine* 24: 1121–1127. <https://doi.org/10.1038/s41591-018-0087-6>.
77. Gross, C., A. Banerjee, D. Tiwari, F. Longo, A.R. White, A.G. Allen, L.M. Schroeder-Carter, J.C. Krzeski, N.A. Elsayed, R. Puckett, E. Klann, R.A. Rivero, S.L. Gourley, and G.J. Bassell. 2019. Isoform-selective phosphoinositide 3-kinase inhibition ameliorates a broad range of fragile X syndrome-associated deficits in a mouse model. *Neuropsychopharmacology* 44: 324–333. <https://doi.org/10.1038/s41386-018-0150-5>.
78. Zogenix. 2019. *Zogenix submits new drug application to U.S. Food & Drug Administration and Marketing authorization application to European Medicines Agency for FINTEPLA® for the treatment of Dravet syndrome – Zogenix, Inc.* <https://zogenixinc.gcs-web.com/news-releases/news-release-details/zogenix-submits-new-drug-application-us-food-drug-administration>. Accessed 7 Jun 2019.
79. Vanhaelen, Q., P. Mamoshina, A.M. Aliper, A. Artemov, K. Lezhnina, I. Ozerov, I. Labat, and A. Zhavoronkov. 2017. Design of efficient computational workflows for in silico drug repurposing. *Drug Discovery Today* 22: 210–222. <https://doi.org/10.1016/j.drudis.2016.09.019>.
80. Ferreira, L.G., and A.D. Andricopulo. 2016. Drug repositioning approaches to parasitic diseases: A medicinal chemistry perspective. *Drug Discovery Today* 21: 1699–1710. <https://doi.org/10.1016/j.drudis.2016.06.021>.
81. Williams, K., E. Bilsland, A. Sparkes, W. Aubrey, M. Young, L.N. Soldatova, K. De Grave, J. Ramon, M. de Clare, W. Sirawaraporn, S.G. Oliver, and R.D. King. 2015. Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. *Journal of the Royal Society, Interface* 12: 20141289–20141289. <https://doi.org/10.1098/rsif.2014.1289>.
82. Alves, V.M., A. Golbraikh, S.J. Capuzzi, K. Liu, W.I. Lam, D.R. Korn, D. Pozefsky, C.H. Andrade, E.N. Muratov, and A. Tropsha. 2018. Multi-descriptor read across (MuDRA): A simple and transparent approach for developing accurate quantitative structure–activity relationship models. *Journal of Chemical Information and Modeling* 58: 1214–1223. <https://doi.org/10.1021/acs.jcim.8b00124>.
83. Ekins, S., A.J. Williams, M.D. Krasowski, and J.S. Freundlich. 2011. In silico repositioning of approved drugs for rare and neglected diseases. *Drug Discovery Today* 16: 298–310. <https://doi.org/10.1016/j.drudis.2011.02.016>.

84. Neves, B.J., R.C. Braga, J.C.B. Bezerra, P.V.L. Cravo, and C.H. Andrade. 2015. In silico repositioning chemogenomics strategy identifies new drugs with potential activity against multiple life stages of *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* 9: e3435. <https://doi.org/10.1371/journal.pntd.0003435>.
85. Govindaraj, R.G., M. Naderi, M. Singha, J. Lemoine, and M. Brylinski. 2018. Large-scale computational drug repositioning to find treatments for rare diseases. *NPJ Systems Biology and Applications* 4: 13. <https://doi.org/10.1038/s41540-018-0050-7>.
86. Sun, P., J. Guo, R. Winnenburg, and J. Baumbach. 2017. Drug repurposing by integrated literature mining and drug–gene–disease triangulation. *Drug Discovery Today* 22: 615–619. <https://doi.org/10.1016/j.drudis.2016.10.008>.
87. Karatzas, E., M.M. Bourdakou, G. Kolios, and G.M. Spyrou. 2017. Drug repurposing in idiopathic pulmonary fibrosis filtered by a bioinformatics-derived composite score. *Scientific Reports* 7: 12569. <https://doi.org/10.1038/s41598-017-12849-8>.