Drug Discovery Today: Therapeutic Strategies
Vol. 8, No. 3–4 2011

Drug repurposing

Cheminformatic/bioinformatic analysis of large corporate databases: Application to drug repurposing

William Loging*, Raul Rodriguez-Esteban, Jon Hill, Tom Freeman, John Miglietta

Computational Biology, Boehringer-Ingelheim Pharmaceuticals Inc., Ridgefield, CT, USA

The approach of drug repositioning is an important consideration for any life science organization. By using knowledge-driven systems in the form of large data stores and applying rational *in silico* experimental design, researchers have generated workflows that are capable of identifying novel uses for drugs that span the therapeutic pipeline and beyond. Both broadly accessible data, such as Medline and Chembank, in addition to internal proprietary data of the company in the form of gene chip experiments, compound screening databases, and clinical trial information play an important role in the success of drug repositioning. By reviewing how current and past successes have been accomplished along with the data used, important stratagems emerge that can provide a wealth of ideas for novel workflows, as well as provide a guide for future discoveries.

Section editor:
Christopher A. Lipinski – Scientific Advisor, Melior Discovery, Waterford, CT 06385-4122, USA

Introduction

As human physiology is comprised of millions of protein–protein interactions it can often be difficult scientifically to predict *de novo* what single or multiple effects a therapeutic molecule may have on disease manifestation or progression. However, either through in-depth scientific review or simple serendipity, the art of drug repositioning has proven itself useful by delivering treatments to alternate indications from those for which the original drug was developed [1–4]. In addressing the declining drug approval rates within the industry, it has been suggested that strategies for repositioning can actually yield more preclinical compounds than similar internal R&D investments [4]. Given this observation and the availability of large data sets to researchers; it is critical that that approaches be in place for investigators to make use of such information.

Through proper experimental design and combining disparate biological, chemical, and clinical data sets that reside in an organization's knowledge warehouse; novel and previously unknown relationships can be found [5–8]. These findings can lead to deeper insights in areas such as disease biology, target/compound selection, and potential toxicities (Fig. 1). There is often no restriction on the type of data used; however their formats can often be an important consideration. For example, data housed in HTML or XML format would require different approaches for mining when compared to data in a relational database or semantic store.

The strategies for mining such large datastores for indication associated linkages can span vast areas of computational science. For the purposes of this review, we will highlight mainstream approaches that have proven themselves useful from the standpoint of delivering quality repositioning opportunities. Such approaches can be broadly broken down into

*Corresponding author:: W. Loging (will.loging@boehringer-ingelheim.com)

 DOI: 10.1016/j.ddstr.2011.06.004

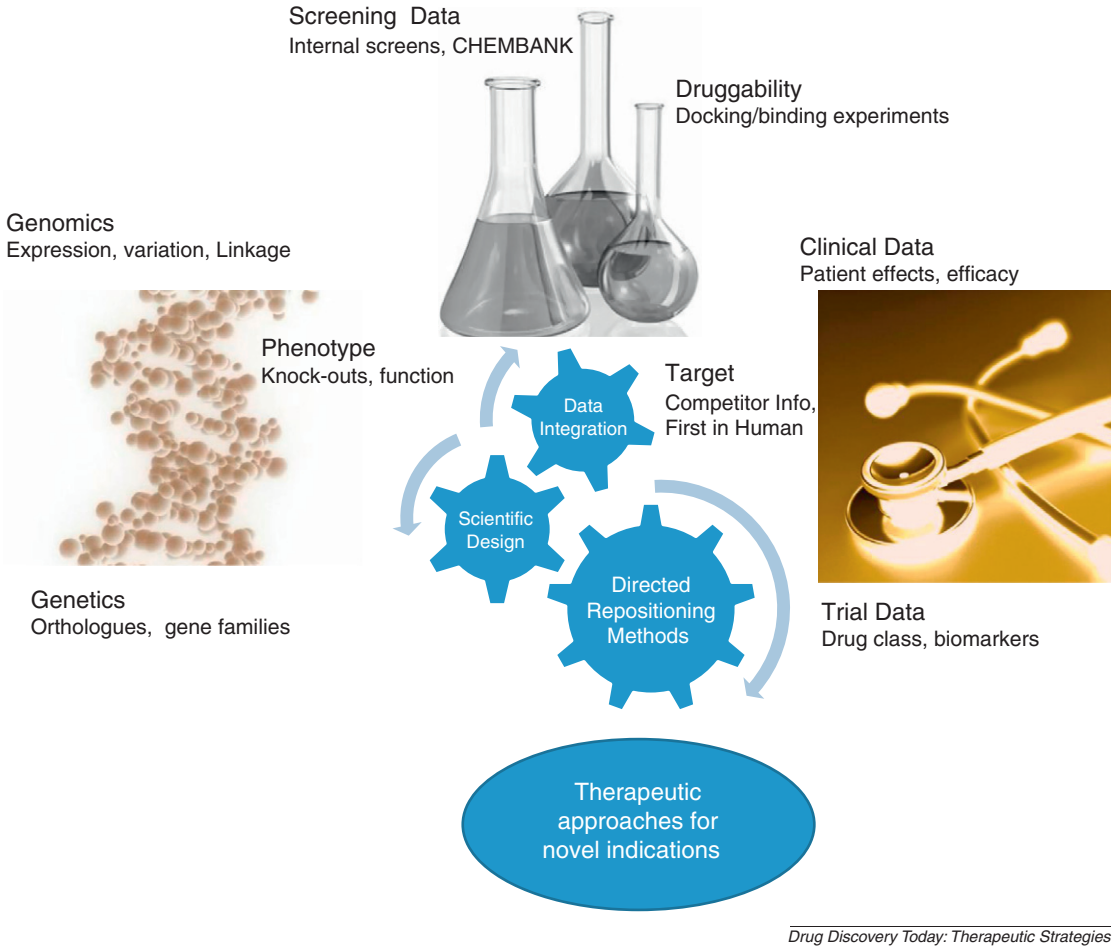

Figure 1. Data sources used in drug repositioning. Computational researchers have access to a wide range of disparate sources of information from across the life sciences. Historically, biological, chemical and clinical data have been often siloed within the drug discovery setting. A critical area in the success of drug repositioning efforts in the integration and application of these diverse data sets, along with rational scientific design. By joining these data domains in a directed fashion, researchers can generate hypotheses for repositioning opportunities.

three categories: those involving biological data (bioinformatics), chemical data (cheminformatics), and text analytics.

Biological data mining

In the generation and capture of biological data, Table 1A lists popular data/information types and associated mining approaches that have revealed repositioning opportunities in the past. Many organizations generate such information either by singleton experiments or through broad focus initiatives. The structure for holding such information can range from simple spreadsheets to advanced semantic data stores. Those companies who are well-positioned to store, process, and display such data to computationally focused researchers will be best able to capitalize on repositioning of both pre-clinical programs, and those under clinical investigation. Even companies without any internally generated data can obtain public source information (Table 1B) and apply data mining strategies to obtain novel linkages and secure intellectual property in this area [2,9].

For those organizations that generate their own biologically derived data to be used for repositioning studies, their approaches can be broken down along disease lines. Generalist organizations do not direct their efforts toward a particular disease; rather, they establish panels of disease-relevant models, and then screen the compounds to be repositioned broadly. Specialists rely on their expertise within a particular disease area to screen relatively larger numbers of compounds against a focused set of diseases. For example, Melior Discovery, generalist in its approach; has established a 'high throughput' *in vivo* pharmacology platform (theraTRACE®) comprising a proprietary multiplexed array of more than 30 *in vivo* models that allows for systematic identification of novel indications for drug candidates at any stage of development [2]. This platform covers therapeutic areas including inflammation, immunology, diabetes and metabolic syndrome, dermatology, cardiovascular, gastrointestinal, psychiatric, neurological and neurodegenerative disorders. Melior stores its data in databases that allow for

Table 1A. Examples of high-level strategies for repositioning and their corresponding mining approaches

Strategy	Mining approach	Example	Pro	Con
In vivo screening	Screen compounds against disease models and derive a phenotyping fingerprint with heatmap visualization	MLR-1023 for diabetes (Melior)	Non-hypothesis driven, very broad in scope, cited 30% true positive return	Requires investment in original data generation
Gene expression screening	Measure gene expression in a panel of relevant tissues, and then perform Bayesian statistical analysis	VVP-808 for diabetes (Verva)	Able to provide a long list of potential opportunities; wealth of hypothesis generation	Additional validation of 'hits' is required; potential long list for follow-up
In vitro screening	Screen compounds for cytotoxicity, then verify approved drugs for clinical efficacy in oncology	Digoxin for oncology (Platz *et al.*)	High-throughput in design, able to generate several candidates for follow-up	More directed toward single cell activity; less likely to obtain tissue related effects
Systematic expert review	Review of existing literature with an eye toward new combinations or diseases	VT-122 for cancer-induced systemic inflammation (Vicus)	Minimal investment in data generation, broad published landscape	Reliant on published knowledge, may require validation of published work
In silico discovery (data integration)	Combine disparate data sources with disease and drug information, then look for 'guilt by association'	Development of DrDKB (Chiang and Butte)	Holistic view of disease landscape from multiple published and external/internal generated data	Additional validation of 'hits' is required; disease data often not available in model organisms
In silico discovery (network based)	Develop network models of biological and pathological conditions to elucidate new useful compound intervention points	Anaxomics networks	Network based analysis, opportunity to identify key proteins nodes that play role in disease	Often limited by visualization and ordering of complex networks

Table 1B. Public source data of use for repositioning

Resource	Data contained	Use to repositioning	URL
Pubmed	Free-text literature abstracts	Highly rich data source for published research. Text-mining or curation may be necessary to integrate with other data sources	http://www.ncbi.nlm.nih.gov/pubmed/
Online Mendelian Inheritance in Man (OMIM)	Fielded free-text descriptions of genes and genetic disorders	Useful for information about human genetic variation and potential phenotypic consequences	http://www.ncbi.nlm.nih.gov/omim
Gene Expression Omnibus	High-throughput gene expression experimental data and study information	Many studies are available from disease tissue samples that can lead to hypothesis generation	http://www.ncbi.nlm.nih.gov/geo/
Mouse Genome Informatics (MGI)	Mouse genetic information	Transgenic mouse phenotypes especially can inform researchers about the potential effects of a therapeutic intervention	http://www.informatics.jax.org/
Kyoto Encyclopedia of Genes and Genomes (KEGG)	Biochemical pathways	A good starting point for construction of disease-relevant networks	http://www.genome.jp/kegg/
BioCarta	Biochemical pathways	Along with KEGG, a good resource for the construction of biological networks	http://www.biocarta.com/
IUPHAR Database	Target and compound information	Excellent source for biology and chemistry information around GPCRs, on channels, and nuclear hormone receptors	http://www.iuphar-db.org
ChEMBL	Bioactive small molecules, including 2D structures and abstracted bioactivities	Can be used as a starting point for construction of biospectra or probing SAR	http://www.ebi.ac.uk/chembldb/
PubChem	Repository of small molecules and biological properties	Another good resource for bioassay and compound structure information	http://pubchem.ncbi.nlm.nih.gov/
ClinicalTrials.gov	Fielded free-text information about clinical research studies	Can provide information about existing repositioning efforts, and drugs that might be available for repositioning	http://clinicaltrials.gov/
SNOMED-CT	Clinical descriptions of diseases and syndromes	An effective starting point for the development of ontologies around disease concepts for text mining	http://www.nlm.nih.gov/research/umls/Snomed/snomed main.html

high throughput mining and 'phenotypic fingerprinting' using heatmap visualization and computational analysis to identify statistically significant changes in dosed animal phenotypes. Coupling such broad phenotypic screening with high throughput mining approaches; led Melior to the repositioning of MLR-1023 for treatment of type 2 diabetes [2,4,10]. MLR-1023 was previously developed for the treatment of gastric ulcers and was halted due to lack of efficacy in a phase 2 trial for that indication [2,10]. As phenotyping led to the discovery of a glucose lowering effect in mice, the primary target was unknown initially and Melior scientists later discovered that Lyn kinase was responsible for the efficacious effect [11]. It is interesting to point out that, had the clinical data for MLR-1023 been analyzed for additional indication effects; glucose lowering would likely have been noted shortly after its trial in the 1980s.

Another example of linking biological data generating platforms to streamlined data mining can be seen by work done at Australia-based Verva. Researchers have developed a diabetes discovery platform (Gene Expression Signature Technology) that uses targeted gene expression measurements from whole cells from a flexible panel of diabetes relevant tissues [4,12]. Using data-driven Bayesian statistical analysis, that is target and mechanism agnostic, within the GES platform; Verva scientists identified VVP808 as a novel non-PPAR insulin sensitizer [12]. VVP808, aka methazolamide, is a carbonic anhydrase inhibitor originally used as a diuretic and in the treatment of glaucoma. The insulin sensitizing action of VVP808 is believed to be independent of its inhibition of carbonic anhydrase.

In addition to mining phenotypic effect databases, other type of data, such as epidemiology information have also been used successfully. Platz *et al.* have incorporated novel biological data by coupling epidemiologic information with an *in vitro* screen of commonly used therapeutic agents to identify novel indications [13]. In contrast to other purely *in silico* efforts (such as that by Chiang and Butte) this is primarily an experimental method. The cytotoxicity of 3187 compounds was determined against two prostrate cancer cell lines. Of the 70 cytotoxic compounds, 38 had regulatory approval and thus a history of clinical use [13]. Twenty of these were known anti-neoplastic agents, while the other 18 were not known as such. When this list of agents was used to interrogate the clinical history of the nearly 48,000 patients in the prospective Health Professionals Follow-up Study, it was determined that there was an inverse risk between use of digoxin, a compound predicted to be anti-neoplastic by the analysis, and prostate cancer risk further strengthening the potential of digoxin as an anti prostrate cancer drug [13].

For strictly *in silico* researchers, their primary way to uncover repositioning opportunities is through new analysis and organization of existing data. For example, Chiang and Butte have created DrDKB, a novel drug information-based research tool, by combining DRUGDX (from The Thomson Corporation) pharmacopeia information, which contains both FDA-approved uses of drugs as well as physician prescribed off-label uses, and a subset of SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) to define and limit the clinical description of diseases and syndromes [14]. This knowledge base was queried via 'guilt-by-association' to identify all disease–drug combinations. If known combinations were excluded, 5500 pairs remained; and when looking at these novel drug uses at clinicaltrials.gov, it was apparent that the suggested novel uses of these drugs was 12 times more likely to be in clinical trials than those not suggested by this analysis [14]. However, one drawback to this method is the large number of individual drug–disease combinations suggested, implying that this method may best be used in combination with an independent method for verification.

Other researchers may organize around a core disease, a core technology, or some combination of the two. For example, scientists at Vicus Therapeutics systematically use published literature in combination with expert knowledge to identify products that comprise novel pairs of generic drugs with established safety profiles that may be of use as adjunctive therapy for oncology [4,9]. Such an approach is also used by scientists at Anaxomics Biotech, however it is not focused on a single therapeutic expertise but used to develop a 'Therapeutic Performance Mapping System' which consists of proprietary patented disease and bio-pathological maps technology and hand-curated databases [15]. Anaxomics curates protein networks that include all molecular entities known to be involved in the disease of interest and create predictive models of the network using pathway analysis tools.

Chemical data mining

The need for drug repositioning can often come after completion of a successful phase 1 clinical trial, demonstrating *in vivo* safety of the candidate drug in healthy human volunteers, and subsequent failure for the intended therapeutic indication in a phase 2 or phase 3 trials in an appropriate patient population. Inclusion of properly selected mechanistic biomarkers should provide evidence that the desired mechanism of action was sufficiently engaged to produce the desired effect. Assuming the drug's mechanism of action is 'hit,' the failure is likely due to a misunderstanding of the role of the mechanism of action in the disease process of interest. The understanding to effectively reposition safe drug entities to appropriate indications lies at the intersection of at least two concepts. First, any given small molecule drug candidate interacts with many proteins based on its properties and concentration and on the local physiological context. Second, complex diseases emerge from the properties of the underlying intra- and intercellular molecular networks and, furthermore, these molecular networks are modular and hierarchical [16]. Information about drug–protein interactions is often incomplete and under appreciated.

However, by generating network-based disease and drug effects landscapes from chemistry-centric data, insights into repositioning opportunities can present themselves.

One such example of a cheminformatics directed repositioning platform can be seen in work by Fliri *et al.* [17]. Large drug–protein activity databases were generated from the interaction of thousands of known drugs and drug-like chemical entities against a section of the druggable proteome. Using heatmap tools for visualization with clustering algorithms – allowed for the creation of a chemical descriptor or 'biospectra' by which small molecules could be grouped based on factors outside their primary structure [18]. It was observed that small molecules fell into distinct clusters which shared specific indication effects. For example, several compounds fell into a cluster marked by those with anti-fungal effects, even though the primary fungal protein target did not exist on the protein panel. Near this antifungal cluster was a cluster of anti-cancer compounds. When the authors looked at the intersection of both clusters, they found compounds like Taxol; a small molecule that exhibits both antifungal and anti-cancer properties [19,20]. Based on the utility of such analyses to identify not only repositioning opportunities, but also group compounds for physiological effects; additional work was performed that linked information of drug–effect relationships which allowed mapping of drugs to physiological processes. Critical to this was computationally linking ever growing corpora of context-specific high quality protein–protein interaction (PPI) data [21]. Weaving data from these areas together has yielded the beginnings of a systems level understanding of drug–molecular network information that is essential not only for drug repurposing but also for successful initial drug purposing and placement. In this specific case, researchers have built context (disease-tissue) specific molecular networks based on PPI into drug–protein and drug–effect relationships that allow inference of which parts of the molecular network are altered by a specific drug and which of these altered subnetworks drive the effects and side effects of that particular drug [22]. Such workflows have also been used to examine the biologically reasons for post-repurposing; such as the example of *nelfinavir* from HIV protease inhibition to oncology [23] (Fig. 2).

In the absence of chemical data generated in house or obtained through vendors, purely *in silico* chemical approaches have been designed. For example, Kinnings *et al.* used a 'chemical systems biology' approach to systematically survey an entire proteome via selective optimization of side activities [24]. Typically, a four step process is utilized: (1) a 3D model of a known drug binding site is obtained; (2) similar, secondary (or off-target) binding sites are identified across the target proteome; (3) drug/off-target interactions are predicted via *in silico* docking methods; (4) the drug is further optimized to enhance potency drug-like properties

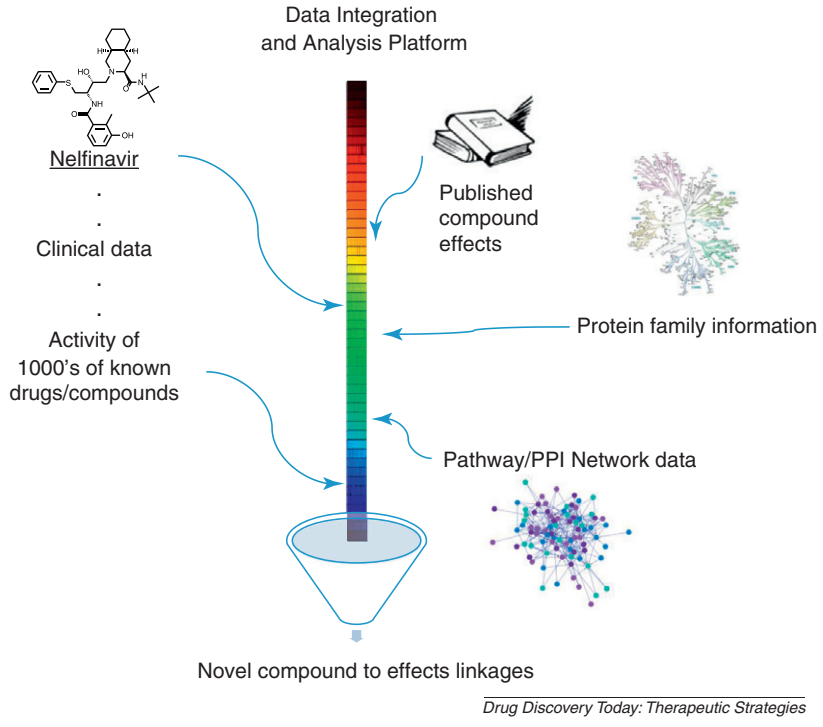

Figure 2. Linkage of compound activity to pathways/prior knowledge. Thousands of compound activity data points are combined in an analysis platform that brings together clinical reports, gene family information (genomics) and Protein-Protein Interaction (PPI) knowledge. Such an approach allows researchers to identify previously unknown connections between small molecules and phenotypic effects. One such example of this is *nelfinavir*, from a primary use of HIV treatment to a possible repositioned use in oncology [23].

considering both primary and secondary targets. Using this approach, Kinnings *et al.* identified known compounds that are hypothesized to be effective in treating multi-drug and extensively drug resistant tuberculosis [24].

Additionally, combinations linking biological screening data to proteins with associated chemical entities have generated hypotheses for new uses. Recently, Buckly *et al.* have described a novel phenotypic screen of 1100 compounds to identify compounds which promote myelination in zebrafish embryos [25]. These compounds were then further characterized by their effect on myelin basic protein mRNA expression. This approach identified novel COX inhibitor chemotypes which have the potential to aid in multiple sclerosis [25].

A slight twist to the traditional approaches of repositioning is to also look for synergistic effects of two known drugs. Using above mentioned data, Iorio *et al.* have described just such an approach [26]. Using transcriptional profiling of drug treated cells they identified similar responses in drug effects and mechanism of action [26]. They then analyzed this data using network theory to uncover unreported effects of known drugs. The network is based on 6100 expression profiles of 5 human cell lines treated with 1309 compounds. One drug identified by this, Fasudil, is hypothesized to be effective against several neurodegenerative diseases [26]. As more single use patents expire for approved drugs, one can expect that approaches to identify known drug synergistic effects will increase.

Text based mining of published knowledge

An important consideration of mining from either the biological or chemical space is how public data and published knowledge are processed. One field that is currently making impacts in the area of drug discovery is that of literature-based discovery [27], which is concerned with the identification of overlooked or 'hidden' relationships between facts in the literature. A natural way to tackle the challenge of repositioning is using text mining to enhance the process of literature-based discovery. Text mining allows performing high-throughput identification of key connections between pieces of evidence scattered in the literature and enables the generation of repositioning hypotheses around drug and disease mechanisms. Recent examples of the utility of text mining can be seen in the areas of identifying a drug's polypharmacology or side effects [28–30].

Even though very few examples of successful repositioning have been described in the text mining literature [31,32], the use of text mining for drug discovery has become routine in the pharmaceutical industry as indicated by the life science products developed by text mining companies such as Ariadne Genomics, Temis, InforSense, and Linguamatics. Such companies offer tools that allow the scanning of literature for potential new indications for drugs, using approaches that range from simple co-occurrence to named entity recognition (NER) or elaborate rule-based natural language processing (NLP). At the same time, companies focused on life science applications such as Genomatix or NextBio have made strides toward adding text mining functionality to their products. However, streamlining and tailoring the discovery process is still an open challenge as the complexities of the task are hard to encapsulate in ready-made software [33]. Moreover, the need for human curation of text mining results limits the level of precision that is acceptable since the scarcest resource is usually the biologists' attention. As with other data-intensive tasks, two common challenges are data visualization and integration of heterogeneous data sources. In the realm of visualization, many approaches have been proposed involving faceted representations of Medline data [34]. Faceted representations allow matrix analysis of information through multiple dimensions, painting the opportunity landscape for target repositioning [35].

For data integration, text mining is routinely used in conjunction with other resources to infer relationships across different domains [35,36]. The heterogeneous datasets (Fig. 3 for sample data amenable to text mining approaches) may cover such spaces as expression data, interaction networks, and clinical trial databases [37], which can be assembled in a multi-dimensional analysis that goes beyond simple association models. The complexity is compounded by the diversity of textual sources, which may include abstracts, full text, patents, clinical trial reports, annotations, etc. The use of semantic web technologies is a new route for integration of such resources [38]. As more widespread use of such tools and semantic approaches are seen, their directed impacts on drug discovery and repositioning will become more prevalent.

The future of repositioning data mining

Much of the value of computational analysis for repositioning comes from the integration of disparate data sources, particularly across the boundaries of the traditional disciplines of clinical medicine, chemistry, biology, and toxicology. Previously, each discipline was heavily siloed – a chemistry database could be rigorous about compound data, but was unlikely to mention the biological target in a meaningful, structured way. As the boundaries between data blur, it is likely that major discoveries remain as large-scale integrations take place. In the more distant future, much of the integration will not be the work of data analysts, but rather will be part of the *design* of the data sources, and will be increasingly transparent to the researcher [8].

In parallel with data integration, there will also be a growing impetus to structure the raw data sources. It is not hard to envision that soon, provision of the data from an article in a means that is appropriate for structured search will be as important for influence and broad citation as the scientific implications of the data itself. As a counteracting force to this, measurements of reliability and providence will grow in application to ensure that analyses based on large, integrated

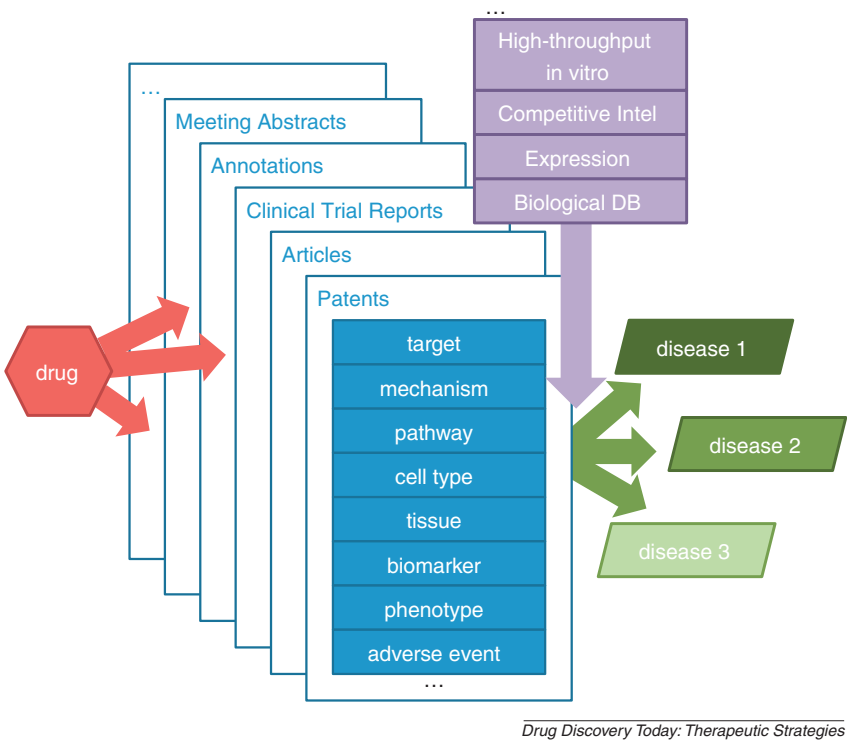

Figure 3. Text sources in literature based mining. Preexisting data in the form of published literature as well as internal organizational documents can be a rich source of information to mine for drug-disease linkages. By linking together printed knowledge from the biological, chemical, and clinical space; researchers can highlight data that supports repositioning hypotheses in a high throughput manner. These computational inquiries typically cover many thousand search combinations and would be too labor intensive if done by traditional Medline exploration.

knowledge bases weigh each parcel of data appropriately. The major areas of biology, chemistry, and clinical data each have the ability to generate repositioning opportunities; however it is when researchers take holistic views of the data and allow for rational scientific design to guide *in silico* experimentation; that one the most potential is noted [39]. Just like wet reagents in a refrigerator are the tools of the *in vitro* investigator, so too are the different datasets that are available to the computational researcher. Best use of experimental design and the tools/data at hand are critical for success in this area.

As repositioning becomes more of a consideration for pharmaceutical research, the nature of how it is conducted will change. Industry pressures and the need to increase productivity are already moving companies to consider multiple diseases to therapeutic entity upstream in the discovery process. Positioning a drug based on its mechanism of action earlier can put companies in better positions to obtain a foothold in new markets from the start, rather than reactively repositioning their drugs as a back-up plan. One such example is Novartis, who created an internal pathway analysis section that advances drug programs based on the pathways that they hit, rather than the diseases they are expected to resolve [7]. This strategy has many scientific merits, as multiple indications can be addressed using a single team focused around the disease causing mechanism. As companies move to work smarter; by making the best use of the data they license and generate; it is envisioned that a new wave of therapeutic products will enter the market. These products will owe their discovery to a new drug discovery paradigm, one in which indications are thought of from holistic, biological disease networks and not as just single targets alone.

References

1 Ashburn, T.T. and Thor, K.B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3, 673–683
2 Campas, C. (2009) Drug repositioning summit: finding new routes to success. *Drug News Perspect.* 22, 126–128
3 Tobinick, E.L. (2009) The value of drug repositioning in the current pharmaceutical market. *Drug News Perspect.* 22, 119–125
4 Sleigh, S.H. and Barton, C.L. (2010) Repurposing strategies for therapeutics. *Pharmaceut. Med.* 24, 151–159
5 Butcher, E.C. *et al.* (2004) Systems biology and drug discovery. *Nat. Biotechnol.* 22, 1253–1259
6 Pache, R. *et al.* (2008) Towards a molecular characterization of pathological pathways. *FEBS Lett.* 582, 1259–1265
7 Chatterjee, S. *et al.* (2009) Identification of novel genes and pathways regulating SREBP transcriptional activity. *PLoS One* 4, e5197
8 Oprea, T.I. *et al.* (2011) Associating drugs, targets and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. *Med. Informatics* 30, 100–111
9 Bascomb, N. *et al.* [Vicus Therapeutics] (2008) Drug combination pharmaceutical compositions and methods for their use. *US20080033027 A1*

10 Lipinski, C.A. *et al.* (1980) Bronchodilator and anti-ulcer phenoxypyrimidinones. *J. Med. Chem.* 23, 1026–1031
11 Reaume, A. and Saporito, M.S. [Melior Pharmaceuticals] (2007) Methods and formulations for modulating lyn kinase activity and treating related disorders. *USP7,776,870*
12 Konstantopoulis, N. *et al.* (2011) A gene expression signature for insulin resistance. *Physiol. Genom.* 43, 110–120
13 Platz, E.A. *et al.* (2011) A novel two-stage, transdisciplinary study identifies digoxin as a possible drug for prostate cancer treatment. *Cancer Discov.* 1, OF66–OF75
14 Chiang, A.P. and Butte, A.J. (2009) Systematic evaluation of drug–disease relationships to identify leads for novel drug uses. *Clin. Pharmacol. Ther.* 86, 507–510
15 Mas Beaventa, J.M. *et al.* [Anaxomics Biotech] (2011) Methods and systems for identifying molecules for processes of biological interest by using knowledge discovery in biological data. *US20110098993 A1*
16 Schadt, E.E. *et al.* (2005) Embracing complexity, inching closer to reality. *Sci. STKE* 295, pe40
17 Fliri, A.F. *et al.* (2009) Drug effects viewed from a signal transduction network perspective. *J. Med. Chem.* 52, 8038–8046
18 Fliri, A.J. *et al.* (2005) Biospectra analysis: model proteome characterizations for linking molecular structure and biological response. *J. Med. Chem.* 48, 6918–6925
19 Fliri, A.J. *et al.* (2005) Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat. Chem. Biol.* 1, 389–397
20 Fliri, A.J. *et al.* (2007) Analysis of system structure-function relationships. *Chem. Med. Chem.* 2, 1774–1782
21 Russell, R.B. and Aloy, P. (2008) Targeting and tinkering with interaction networks. *Nat. Chem. Biol.* 4, 666–673
22 Fliri, A.J. *et al.* (2010) Cause-effect relationships in medicine: a protein network perspective. *Trends Pharm. Sci.* 31, 547–555
23 Xie, L. *et al.* (2011) Drug discovery using chemical systems biology: weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir. *PLoS Comput. Biol.* 7, e1002037
24 Kinnings, S.L. *et al.* (2009) Drug discovery using chemical systems biology: repositioning the safe medicine comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput. Biol.* 5, e1000423
25 Buckley, C.E. *et al.* (2010) Drug reprofiling using zebrafish identifies novel compounds with potential pro-myelination effects. *Neuropharmacology* 59, 149–159
26 Iorio, F. *et al.* (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14621–14626
27 Weeber, M. (2007) Drug discovery as an example of literature-based discovery. In *Computational Discovery of Scientific Knowledge* (Džeroski, S. and Todorovski, L., eds), pp. 290–306, Springer
28 Agarwal, P. and Searls, D.B. (2008) Literature mining in support of drug discovery. *Brief. Bioinform.* 9, 479–492
29 Frijters, R. *et al.* (2010) Literature mining for the discovery of hidden connections between drugs, genes and diseases. *PLoS Comput. Biol.* 23, e1000943
30 Plake, C. and Schroeder, M. (2011) Computational polypharmacology with text mining and ontologies. *Curr. Pharm. Biotechnol.* 12, 449–457
31 Deftereos, S.N. *et al.* (2011) Drug repurposing and adverse event prediction using high-throughput literature analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3, 323–334
32 Shaughnessy, A.F. (2011) Old drugs, new tricks. *Br. Med. J.* 342, d741
33 Rodriguez-Esteban, R. (2009) Biomedical text mining and its applications. *PLoS Comput. Biol.* 5, e1000597
34 Lu, Z. (2011) PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)* 2011, baq036
35 Campbell, S.J. *et al.* (2010) Visualizing the drug target landscape. *Drug Discov. Today* 15, 3–15
36 von Eichborn, J. *et al.* (2011) PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res.* 39 (Database issue), D1060–D1066
37 Dudley, J.T. *et al.* (2010) Drug discovery in a multidimensional world: systems, patterns, and networks. *J. Cardiovasc. Transl. Res.* 3, 438–447
38 Qu, X.A. *et al.* (2009) Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships. *BMC Bioinform.* 10 (Suppl. 5), S4
39 Schadt, E.E. *et al.* (2009) A network view of disease and compound screening. *Nat. Rev. Drug Discov.* 8, 286–295