SciBite
an ELSEVIER company

Whitepaper

Semantic Analytics: A systematic, data-driven approach to drug repositioning

SciBite
an ELSEVIER company

Semantic Analytics: A systematic, data-driven approach to drug repositioning

The past two decades have seen an increase in drug resistance rates in cancer and chronic disease and a rise in multi-drug resistant infectious diseases. Coupled with continued population growth, there is a clear and urgent need for new and effective therapeutics. Following years of significant growth, Pharmaceutical companies are witnessing reduced sales growth rates and are facing increasing cost pressures from both payers and competitors[1]. Adding to this, the average cost of de novo drug development is now in excess of $2.5 billion and continues to grow[2]. While there has been an upturn in drug approvals in recent years[3], productivity is not keeping up with increasing R&D costs.

Due to rising development costs, owing in no small part to a continued reliance on costly high throughput in vitro screening, a growing number of biotechs and pharmaceutical companies are turning to Drug Repositioning Repurposing and Rescue (hereafter referred to simply as 'Drug Repositioning'). The reasoning behind this is to improve the efficiency of the development process and maximise the value of their R&D investments by focussing on drugs that have near-term clinical potential. The goal of Drug Repositioning is to discover new uses for drugs to treat clinical indications other than those for which they were originally intended[4].

Drug Repositioning can also be applied to molecules that are in the development pipeline or to rescue assets that underwent some incomplete initial development and have a proven safety profile but were mothballed for business reasons, such as a change in therapeutic focus, yet remain an important asset to the company.

For example AZT was originally developed with the goal of treating cancer, but was repositioned and approved for use as a treatment for HIV[5].

Targeted re-use of drugs can bring new therapies to market in approximately half the budget and time required by the traditional drug development cycle[6]. This is typically because candidates for repositioning have pre-existing efficacy, pharmacokinetic, pharmacodynamic, toxicity and dosing data, often coupled with well characterised biological knowledge – all of which is costly and time-consuming to obtain for a new molecular entity (NME)[7].

The risks associated with development can also be significantly reduced due to a higher probability of success and reduced attrition – 25% of phase 2 candidates succeed to approval, compared with just 10% for a NME[8].

1 For example see: http://www.ey.com/Publication/vwLUAssets/EY-beyond-borders-2016/$FILE/EY-beyond-borders-2016.pdf and https://www.statista.com/topics/1764/global-pharmaceutical-industry/

2 Dimasi J. A., Grabowski H. G., & Hansen R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics, 47, 20-33 Dimasi J. A., Grabowski H. G., & Hansen R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics, 47, 20-33.

3 For example, see: http://cmr.clarivate.com/pdf/2016_CMR_Executive_Summary.pdf

4 Ashburn T.T. and Thor K.B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. Nature Reviews Drug Discovery 3:673-683.

5 Strahl C., Blackburn E. H. (1996). Effects of reverse transcriptase inhibitors on telomere length and telomerase activity in two immortalised human cell lines. Mol. Cell. Biol. 16:53–65.

6 Elvidge, S. (2016). Getting the drug repositioning genie out of the bottle. Life Sci. Leader 14-18; Novac N., (2013). Challenges and opportunities of drug repositioning. Trends Pharmacol. Sci. 34:267-272; Smith, R.B. (2012). Repositioned drugs: integrating intellectual property and regulatory strategies. Drug Discovery Today: Therap. Strat. 8(3):131-137.

7 Tobinick E.L. (2009). The value of drug repositioning in the current pharmaceutical market. Drug news & perspectives 22:119-125.

8 Barratt M.J., Frail D.E. (2012). Drug Repositioning: Bringing New Life to Shelved Assets and Existing Drugs. Hoboken, NJ: John Wiley & Sons.

In 2013, new indications for existing drugs accounted for 20% of the products introduced[9]. Drug Repositioning also has the potential to extend patent exclusivity and raises the possibility to outlicense rights to the new indication[10].

The past 20 years have also seen increasing interest in rare diseases – defined as diseases that affect fewer 1:1500 people in the US. Even though the use of Next Generation Sequencing (NGS)-based profiling of patients to identify underlying genetic variations has resulted in greater numbers of people being diagnosed with rare diseases, the R&D costs associated with developing a de novo therapy make it intractable to develop new drugs for rare diseases. Because of this, Drug Repositioning has received attention and support at the national government level in many countries, including economic incentives and streamlined regulatory pathways to encourage pharma to invest in the development of rare disease therapies, such as the US FDA Orphan Drug Act (ODA) and legislation in Japan, Australia and Europe[11]. These factors have led to the strategy of 'indication hopping' – initially targeting a disease area that attracts financial incentives then, once approval is gained, repurpose it for related indications and increasing the overall value of the drug[12].

Hence, Drug Repositioning has the potential to provide a better return on investment than an NME and has the potential to provide patients with access to better therapeutics, with limited side effects in a more timely manner and deliver treatments that are more affordable for large, otherwise unserved patient communities[13].

There have been several high profile and highly profitable examples of repositioned drugs. For example, sildenafil was originally developed as a treatment for angina but the resulting side effects revealed an opportunity to reposition it for male erectile disfunction. It is now marketed by Pfizer as the blockbuster drug Viagra, in addition to its subsequent approval for pulmonary hypertension[14]. Finasteride was developed to treat benign prostatic hyperplasia, but was observed to have a side effect of hair growth and is now used as a treatment for male pattern baldness[15]. However, these examples are a result of serendipity rather than design.

Testing all drugs against all targets experimentally is currently infeasible due to cost and technical constraints[16]. A more rational approach to Drug Repositioning is to apply the knowledge of one drug to better characterise another. The exponential growth of data from 'omics' initiatives has resulted in detailed biological pathway maps, drug profiles and disease phenotypes. These, coupled with the availability of increasingly powerful computational resources to identify patterns and relationships, has made computational approaches to repositioning a viable proposition[17]. For example, in silico repositioning has highlighted tricyclic antidepressants as potential inhibitors of small lung cancer[18] and a possible use of the anticonvulsant Topiramate to treat irritable bowel disease[19].

However, the majority of computational repositioning initiatives have been limited to one or two types of data, such as identifying drugs with similar chemical structures or drugs that have similar gene expression profiles.

9 Graul A.I., Cruces E. and Stringer M. (2014). The year's new drugs and biologics, 2013: Part I, Drugs Today (Barc) 50(1):51-100; Rodriguez-Esteban R.A. (2016). A drug-centric view of drug development: how drugs spread from disease to disease. PLoS Comput. Biol. 12:e1004852.

10 Mucke H.A.M. and Mucke E. (2015). Sources and targets for drug repurposing: landscaping transitions in therapeutic space. Drug Repurps. Rescue Reposition 1: 22-27 ; Vanhaelen Q., Mamoshina P., Aliper AM., Artemov A., Lezhnina K., Ozerov I., Labat I. and Zhavoronkov A. (2016). Design of efficient computational workflows for in silico drug repurposing. Drug Discovery Today 22:210-222.

11 Tambuyzer E. (2010). Rare diseases, orphan drugs and their regulation: questions and misconceptions. Nat. Rev. Drug Discov. 9:921-929; Melnikova I. (2012). Rare diseases and orphan drugs. Nat. Rev. Drug Discov. 11(4):267-8.

12 Shelley S. (2016). http://pharmaceuticalcommerce.com/brand-marketing-communications/the-business-of-orphan-drugs-is-booming/. Published on August 26, 2015. Accessed on August 23rd, 2017.

13 Deotarse P., Jain A., Baile M., Kohle N. and Kulkarni A. (2015). Drug repositioning: A review. International Journal of Pharma Research & Review 4:51-58; Njogu P.M. and Chibale K. (2013). Recent developments in rationally designed multi-target antiprotozoan agents. Curr. Med. Chem. 20(13):1715-1742; Tartaglia L.A. (2006). Complementary new approaches enable repositioning of failed drug candidates. Exp. Opin. Investiga. Drugs 15(11):1295-1298.

14 Ghofrani H.A., Osterloh I.H. and Grimminger F. (2006). Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond Nat. Rev. Drug Discov. 5(8):689-702.

15 Gormley G.J., Stoner E., Bruskewitz R.C, Imperato-McGinley J., Walsh P.C., McConnell J.D., Andriole G.L., Geller J., Bracken B.R., Tenover J.S., et al. (1992) The effect of finasteride in men with benign prostatic hyperplasia. The Finasteride Study Group. N. Engl. J. Med. 327:1185-1191.

16 Li Y.Y., An J. and Jones S.J.M. (2011). A computational approach to finding novel targets for existing drugs. PLoS Computational Biology 7:e1002139.

17 Wu Z., et al. (2013). Network-based drug repositioning. Mol. BioSyst. 9:1268-1281; Zou J., et al. (2013). Advanced systems biology methods in drug discovery and translational biomedicine. BioMed Res. Int. 742835.

18 Jahchan N.S., Dudley J.T., Mazur P.K., Flores N., Yang D., Palmerton A., et al. (2013). A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. Cancer Discov. 3(12):1364-1377.

19 Dudley J.T., Sirota M., Shenoy M., Pai R.K., Roedder S., Chiang A.P., Morgan A.A., Sarwal M.M., Pasricha P.J. and Butte (2011). Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Science Trans. Med. 3:96ra76.

SciBite
an ELSEVIER company

This whitepaper describes how Semantic Analytics enables a scalable, integrated approach to Drug Repositioning, releases the potential of and will enable Pharmaceutical companies to:

- Unlock the wealth of diverse unstructured qualitative biomedical content
- Integrate data across silos and across data types
- Develop a deep and rich understanding of their drugs, their targets and their interactions

... and ultimately to undertake more effective Drug Repositioning initiatives, faster and at a lower cost.

Enabling an integrative, data-driven approach to drug repositioning

To be effective, a Drug Repositioning strategy must be built on several critical foundations. Namely the ability to:

- Incorporate a range of diverse data types
- Unlock the potential of unstructured, qualitative scientific content including publications, patents and the content of electronic laboratory notebooks
- Contextualise data and make connections across disparate data sources

Each of these is described below.

Incorporating a range of data sources

As outlined below, several types of data, including both drug and disease-based data, can be used to inform Drug Repositioning strategies[20].

Structural Data

Many drugs bind multiple targets and therefore affect multiple pathways. This phenomenon is referred to as drug promiscuity, which can in turn lead to the identification of secondary indications for a drug[21]. The theory that drugs with similar chemical structures can have similar biological activities (the Similar Property Principle) is the basis for virtual screens to identify drugs with known indications that are structurally similar to a drug of interest, and then apply the original drug of interest to those indications[22]. Likewise, molecular docking and virtual screens based on drug target binding site similarity can be used to identify secondary targets involved in other indications, thus identifying new repositioning opportunities[23].

Screens based on drug and target binding site similarity require reliable 2D and 3D molecular structures, which are not always available[24], for example there are experimental limitations to obtaining full 3D structural data for membrane-bound proteins[25]. When using chemical structural data in this way, it's important to use the appropriate active form of a drug for analysis. For example, many drugs, such as tamoxifen, are administered as pro-drugs or undergo uncharacterized chemical transformations after metabolism to achieve their active form[26].

Genomic Data

The fact that many genes are pleiotropic in nature and influence two or more seemingly unrelated disease indications is the basis of genomic-based approaches to identifying repositioning opportunities, such as comparing the gene expression profiles derived from two drugs. Several resources exist to aid genomic-based Drug Repositioning strategies, including Connectivity Map (CMap[27]) which includes thousands of gene expression profiles induced by a wide range of FDA approved drugs, and the Library of Integrated Network-Based Cellular Signatures (LINCS[28]).

20 For detailed reviews see Hodos R.A., Kidd B.A., Shameer J., Readhead B.P. and Dudley J.T. (2016). In silico methods for drug repurposing and pharmacology. WIREs Syst Biol Med. 8(3):186-210; and also Li J., Zheng S., Chen B., Butte A.J., Swamidass S.J. and Lu Z. (2016). A Survey of Current Trends in Computational Drug Repositioning. Briefings in Bioinformatics 17(1): 2-12.

21 Haupt V.J., Daminelli S. and Schroeder M. (2013). Drug discovery: predicting promiscuity. Nature 462:167-168; Dudley J.T., Sirota M., Shenoy M., Pai R.K., Roedder S., Chiang A.P., Morgan A.A., Sarwal M.M., Pasricha P.J. and Butte (2011). Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Science Trans. Med. 3:96ra76.

22 Keiser M.J., Setola V., Irwin J.J., Laggner C., Abbas A.I., Hufeisen S.J., Jensen N.H., et al. (2009). Predicting new molecular targets for known drugs. Nature 462(7270):175-181; Sanseau P. and Koehler J. (2011). Editorial: Computational methods for drug repurposing. Briefings in Bioinformatics 12(4):301-302.

23 Moriaud F., Richard S.B., Adcock S.A., Chanas-Martin L., Surgand J.-S., Jelloul M.B. and Delfaud F. (2011). Identifying drug repositioning candidates by mining the protein data bank. Briefings in Bioinformatics 12(4):336-340; Li Y.Y., An J. and Jones S.J.M. (2011). A computational approach to finding novel targets for existing drugs. PLoS Computational Biology 7:e1002139.

24 Vanhaelen Q., Mamoshina P., Aliper AM., Artemov A., Lezhnina K., Ozerov I., Labat I. and Zhavoronkov A. (2016). Design of efficient computational workflows for in silico drug repurposing. Drug Discovery Today 22:210-222.

25 Hodos R.A., Kidd B.A., Shameer J., Readhead B.P. and Dudley J.T. (2016). In silico methods for drug repurposing and pharmacology. WIREs Syst Biol Med. 8(3):186-210.

26 Furr B.J.A. and Jordan V.C. (1984). The pharmacology and clinical uses of tamoxifen. Pharmacol. Ther. 25:127-205.

27 Lamb J., Crawford E.D., Peck D., Modell J.W., Blat I.C., et al. (2006). The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313(5795): 1929-1935.

28 Vidovic D., Koleti A. and Schurer S.C. (2014). Large-scale integration of small molecule-induced genome-wide transcriptional responses, Kinome-wide binding affinities and cell-growth inhibition profiles reveal global trends characterizing system-level drug action. Front. Genet. 5:342.

CMap and LINCS can be used to identify a 'signature' – a pattern of genes that are up- and down-regulated in a given disease which can be used to identify other diseases that induce similar patterns. The theory behind the so-called guilt-by-association (GBA) method is that drugs used to treat one disease may be repositioned to treat diseases with a similar signature[29]. Similarly, signature inversion is based on the rationale that drugs that induce a particular gene expression signature may be used to treat diseases that have the opposite signature[30].

However, this gives no consideration of underlying mechanism and care must be taken to ensure that expression profile associated with a disease is not a consequence of the disease state rather than its cause. In addition, diseases that involve multiple tissues and/or organ systems may not be amenable to approaches that rely on the genomic expression profiles of isolated cell lines.

Disease and Phenotypic Data

Disease-based approaches are based on the premise that if two diseases share multiple approved drugs, any drugs approved for only one of the diseases could potentially be used as a treatment for the other disease[31].

Phenotypic-based approaches to Drug Repositioning focus on disease commonalities between diseases, such as co-morbidities, and analysis of side effect similarities. Side effects normally have negative connotations, but analysis of the side effects associated with a drug, such as those found in the Side Effect Resource (SIDER), can suggest possible new therapeutic areas for repositioning. For example, drugs with similar side effect profiles to known drugs or drug classes, may have a common mechanism of action[32].

This approach can be limited by the fact that the full side effect profile of a drug is unlikely to be known until after many years of post-market surveillance and the fact that the same side effect can occur for different reasons.

However, while it would appear to be 'short cutting' target and mechanism of action information, there are cases where this approach has been shown to be more predictive of repositioning opportunities than relying on chemical structure or target information[33].

Many diseases are regulated by multiple signalling pathways, so targeting one pathway is insufficient to successfully treat the disease. Drug combination strategies to treat such diseases often arise from clinical experience and it is not practical to consider all possible drug combinations when searching for repositioning opportunities[34]. However, the search can be limited to drugs that can be co-prescribed as they usually don't share side effects, particularly if used in combination with genomic profiles to predict effective combinations[35].

Unlocking unstructured information from the biomedical literature

A plethora of information relating to associations between genes, targets and diseases can be found in unstructured biomedical sources, such as PubMed, RightFind™, ClinicalTrials.gov and patent databases. However, the vast majority of these sources are not designed with searching in mind. Compounding these issues, manual curation can't keep up with the amount of data in the literature[36] – curated information is often not available for several months after publication and, even then, is often only accessible via subscription databases. Computer-based text analytics offers the possibility to access and use the content biomedical literature more effectively and capture data that may have been missed by manual curation[37]. For example, text analytics has been used to analyse the unstructured text within drug labels to infer whether two drugs share the same target[38]. However, many computational approaches struggle to deal with the complexity and variability of unstructured scientific language.

29 Iorio F., Rittman T., Ge H., Menden M. and Saez-Rodriguez J. (2013). Transcriptional data: a new gateway to drug repositioning? Drug Discovery Today 18:350-357.

30 Dudley J.T., Sirota M., Shenoy M., Pai R.K., Roedder S., Chiang A.P., Morgan A.A., Sarwal M.M., Pasricha P.J. and Butte (2011). Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Science Trans. Med. 3:96ra76.

31 Chiang A.P., Butte A.J. (2009) Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. Clin. Pharmacol Ther. 86:507-510.

32 Yang L. and Agarwal P. (2011). Systematic drug repositioning based on clinical side-effects. PLos ONE 6(12):e28025.

33 Zhang P., Wang F., Hu J. and Sorrentino R. (2013) Exploring the relationship between drug side-effects and therapeutic indications. AMIA Annu. Symp. Proc. 1568-1577.

34 Huang L., Li F., Sheng J., Xia X., Ma J., Zhan M. and Wong S.T.C. (2014). Drugcomboranker: drug combination discovery based on target network analysis. Bioinformatics 30:i228-i236.

35 Sun Y., Xiong Y., Xu Q. and Wei D. (2014). A hadoop-based method to predict potential effective drug combination. BioMed Research International 2014:196858.

36 Andronis C., Sharma A., Virvilis V., Deftereos S. and Persidis A. (2011). Literature mining, ontologies and information visualization for drug repurposing. Briefings in Bioinformatics 12:357-368.

37 Liu, R.L. and Shih C.C. (2014). Identification of highly related references about gene-disease association. BMC Bioinformatics 15:286; Bravo A., Piñero J., Queralt-Rosinach N.A., Rautschka M. and Furlong L.I. (2015). Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. BMC Bioinformatics 16(1):55+.

38 Campillos M., Kuhn M., Gavin A.C., Jensen L.J., Bork P. (2008). Drug target identification using side-effect similarity. Science 321:263–266.

Structured domain knowledge, in the form of high quality biomedical vocabularies and ontologies such as ChEMBL (bioactive molecules) and MeSH (disease classifications) provides a common language and framework for mapping biological relationships. It can range from simple lists containing all of the known terms for the same real world "thing" to hierarchical groupings and classifications of scientifically-related concepts. But no single vocabulary or ontology is comprehensive in either depth or breadth. For example, no single vocabulary encompasses every possible term or phrase used for every indication. Biomedical vocabularies and ontologies often have different purposes and can be mutually inconsistent, with the result that they are poorly integrated with each other[39]. An integrative, data-driven approach to Drug Repositioning requires the integration of multiple complementary ontologies, enabling the identification of all possible uses and variants of a term of interest.

Semantic analytics: contextualizing and linking data

Whilst several initiatives have integrated data from multiple sources to create secondary data sources, such as DisGeNET[40] and Malacards[41], there isn't a single definitive source of data for Drug Repositioning, and the trend is for information to be unstructured and increasingly decentralized, with new sources of valuable data arising every year.

While each of the individual data types and sources described above can yield valuable clues for Drug Repositioning, using multiple data sources can give a clearer picture of a drug and its effects, and lessen the potential noise or bias of any one source. Some repositioning initiatives have combined structural data with other types of data, e.g. chemical similarity plus side effect data or drug-target interactions and gene similarity data[42]. However, rather than limit a search to one or two types, such as drug-disease or drug-gene interactions, the ability to combine drug, target and disease information provides the basis for a more holistic approach and greater confidence in the results[43]. For example, it is possible to infer a drug-disease association based on linking genotype-disease and drug-gene information[44]. However, achieving a holistic approach has been hindered by the challenges associated with integrating heterogeneous measurements and/or diverse data sources[45].

The vocabularies and ontologies described in the previous section are the foundation for Semantic Analytics, which applies an explicit, unique meaning and description to a term. The application of Semantic Analytics to unstructured text enables it to be contextualised, understood and used as high quality, actionable data, irrespective of its source. The ability to uniformly contextualise data by semantic enrichment provides the basis for linking data from heterogeneous biomedical sources. It facilitates the generation of a network, or semantic model, of interconnected facts which provides a convenient way to represent the integrated data necessary to identify repositioning opportunities[46]. For example, disease-specific drug-protein connectivity maps can be generated by integrating protein interaction data with literature mining. Based on this approach, diltiazem (an anti-hypertensive agent) and quinidine (an anti-arrhythmia agent) were proposed as candidate treatments for Alzheimer's disease[47].

In addition, there has been growing interest in applying machine learning techniques to Drug Repositioning, by identifying patterns indicative of a drug interacting with a target and use classification and learning to predict new indications for the drug[48].

39 Mullen, J., Cockell S.J., Woollard P. and Wipat A. (2016). An integrated data driven approach to drug repositioning using gene-disease associations. PLoS One 11(5):e0155811.

40 Bauer-Mehren A., Rautschka M., Sanz F. and Furlong L.I. (2010). DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics 26(22):2924-2926.

41 Rappaport N., Twik M., Nativ N., Stelzer G., Bahir I., Stein T.I., et al. (2014). MalaCards: A comprehensive automatically-mined database of human diseases. Current protocols in bioinformatics/editorial board, Andreas D. Baxevanis et al. 47:1.24-1.24.19.

42 For example see Wang Y., Chen S., Deng N., et al. (2013). Drug repositioning by kernel-based integration of molecular structure, molecular activity, and phenotype data. PLoS One 8(11):e78518; Tan F.J., Yang R.Z., Deng N., et al. (2014). Drug repositioning by applying 'expression profiles' generated by integrating chemical structure similarity and gene semantic similarity. Mol. Biosys. 10(5):1126-1138.

43 Chen B., Ding Y. and Wild D.J. (2012). Assessing drug target association using semantic linked data. PLOS computational biology 8(7):e1002574; He B., Tang J., Ding J., Wang H., Sun Y., Shin J.H. et al. (2011). Mining relational paths in integrated biomedical data. PLoS One 6(12):e27506+; Lee H.S. et al. (2012). Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug. BMC Syst. Biol. 6:80.

44 Hoehndorf R., Oellrich A., Rebholz-Schuhmann D., et al. (2012). Linking PharmGKB to phenotype studies and animal models of disease for drug repurposing. Pac. Symp. Biocomput. 388-399.

45 Li J., Zheng S., Chen B., Butte A.J., Swamidass S.J. and Lu Z. (2016). A Survey of Current Trends in Computational Drug Repositioning. Briefings in Bioinformatics 17(1): 2-12.

46 Betzler N., van Bevern R., Fellows M.R., Komusiewicz C. and Niedermeier R. (2011). Parameterized algorithmics for finding connected motifs in biological networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics/IEEE, ACM 8(5):1296-1308.

47 Li J., Zhu X., Chen J.Y. (2009). Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. PLoS Comput. Biol. 5:7 e1000450..

48 For example, see: Napolitano F., Zhao Y., Moreira V.M., Tagliaferri R., Kere J., D'Amato M. and Greco D. (2013). Drug repositioning: A machine-learning approach through data integration. J. Cheminformatics 5:30.

Despite significant advances in the technology, sophisticated tools such as machine learning ultimately rely on good data. Semantic enrichment provides the contextualised data necessary to enable machine learning to be effective.

Semantic Analytics enables the detection of relevant information within unstructured biomedical literature enabling text to be treated as data. As the examples in the following section illustrate, leveraging biomedical literature in this way provides the rich, interconnected evidence necessary to support a data-driven approach to Drug Repositioning via the accumulation, integration and analysis of a disparate set of data sources and data types.

Applying semantic analytics to drug repositioning

The ability to contextualize unstructured biomedical literature and to treat the resulting data as an integral part of a Drug Repositioning strategy is fundamental to achieving the holistic view necessary to identify all potential repositioning opportunities and make well informed decisions. This section provides examples of how SciBite's Platform enables a systematic, data-driven approach to Drug Repositioning, including the ability to rapidly review of all available information via user-focussed scientific browsing and also how it forms the basis of a more automated, large scale analyses.

User-focused scientific browsing to validate potential repositioning opportunities

Drug repositioning initiatives that focus on one or two data types can yield long lists of potential indications for repositioning, each of which requires further investigation. Validating some of the options may be possible by connecting people with relevant knowledge, such as experts on a given drug target and experts in each disease, but a more comprehensive evaluation typically requires a lengthy literature search. However, relying on full text searches of biomedical resources such as PubMed and patents to reveal important relationships or trends is risky and time consuming due to the siloed disconnected nature of these sources.

Semantic Analytics can highlight areas for focus. For example, a virtual screen based on chemical structure of a drug of interest can identify similar drugs associated with other indications. The resulting list of indications can be used to mine a semantic network to identify drug targets associated with i) each of the structurally similar drugs and/or ii) each of the associated indications. Indications associated with one or more targets in common with the drug of interest can be prioritized as being the most likely initial candidates for further investigation.
The interconnected nature of data within a semantic network makes it simple to collate all known information about a disease of interest, thus increasing the amount of evidence available to resulting in a more validated line of enquiry. Rather than relying on static, manually curated information, the knowledge encapsulated in SciBite's semantic network grows over time.

Connecting additional data sources containing information such as patent life, toxicity, mode of action, route of delivery and other chemical or disease-specific information can provide valuable input into the prioritization of Drug Repositioning opportunities[49] and can help rule out things that don't make sense. For example, drugs that require specialist storage and testing requirements would not be realistic candidates for repositioning to treat infectious diseases in more remote parts of the world. Similarly, the inclusion of drug approval information can identify indications where a drug was approved in some regions but not others.

Coupled with information about repositioning successes and failures, a growing corpus of connected information can also be used to identify the characteristics associated with repositionable drugs and indications, which can help refine future repositioning initiatives.

Comprehensive semantic analytics to identify new repositioning opportunities

The benefits of applying Semantic Analytics to Drug Repositioning activities can be illustrated using the drug Rivastigmine as a real-world example. A traditional approach to repositioning would involve the following high level activities:

1. A search of DailyMed results in over 20 labels. Reading each of these reveals that Rivastigmine is used as a treatment for dementia associated with Alzheimer's and Parkinson's disease
2. A review of the ChEMBL record for the mechanism of action of Rivastigmine reveals that it is an inhibitor of acetylcholinesterase (ACHE) and butyrylcholinesterase (BCHE).
3. Individual searches of DisGeNET for both ACHE and BCHE generates a list of over 40 indications associated with these two protein targets
4. Reducing the list to those indications that have the most similar phenotypes to the on-label drug indications from DailyMed involves a thorough review of the information associated with each of the 40 indications associated with ACHE and BCHE

49 Andronis C., Sharma A., Virvilis V. et al. (2011). Literature mining, ontologies and information visualization for drug repurposing. Brief. Bioinform. 12(4):357-368.

This is a laborious manual process which is reliant on human judgement and therefore susceptible to human fallibility. Reviewing reams of information over several days increases the likelihood of the reviewer missing one or more of the linkages.

Using Semantic Analytics associations between drugs, targets and indications can generate a knowledge network which integrates data from different sources. For example, SciBite's platform integrates structured sources with unstructured biomedical literature into network analyses, providing a much richer picture and revealing more associations than relying on structured data alone.

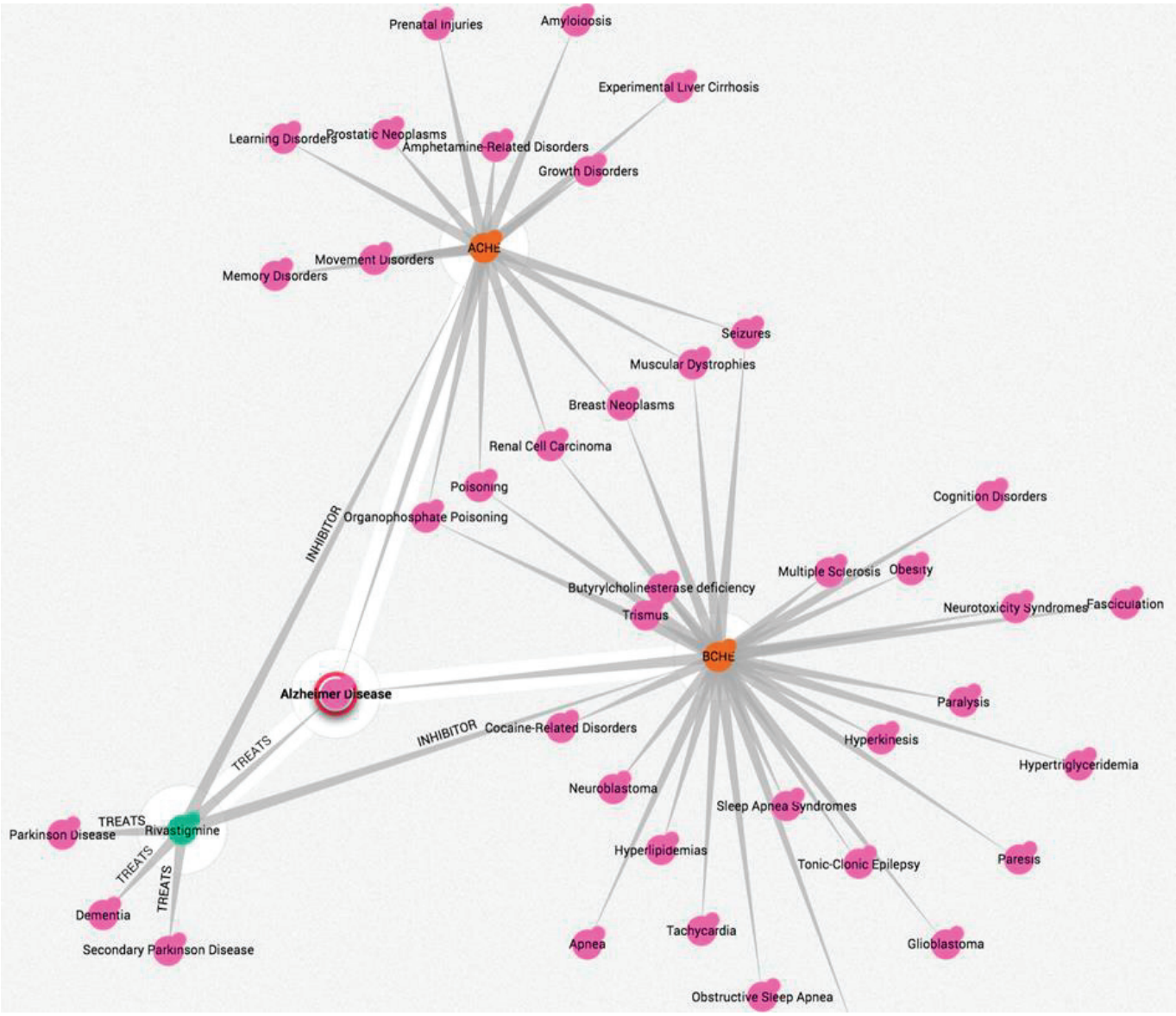

Figure 5: A semantic network generated by the SciBite platform and visualised using Linkurious[50] illustrating relationships between the Drug Rivastigmine (green node), drug targets from ChEMBL associated with Rivastigmine (ACHE and BCHE; orange nodes) and all indications associated with these drug targets from DailyMed and DisGeNET (pink nodes).

50 https://linkurio.us

Rather than read the literature on each of the indications, applications such as SciBite's TERMite[51] can be used to extract disease-phenotype information from the biomedical literature. By applying a filter for those indications with similar phenotype profiles to the on-label drug indications[52], it is possible to rapidly reduce the number of potential indications. Specifically, such analysis reveals Multiple Sclerosis (MS) as a candidate for repositioning of Rivastigmine based on a mechanism of action associated with the target BCHE and a phenotypic similarity to Alzheimer's Disease. Phenotype connections that were relevant to this analysis were mainly those associated with cognitive impairment, recognised as a frequent symptom with MS and, more obviously, with AD.

A review of DrugBank doesn't provide any evidence for a link between Rivastigmine and MS. Indeed, a search of the biomedical literature for publications mentioning both Rivastigmine and any disease indication ranks MS low on the list.

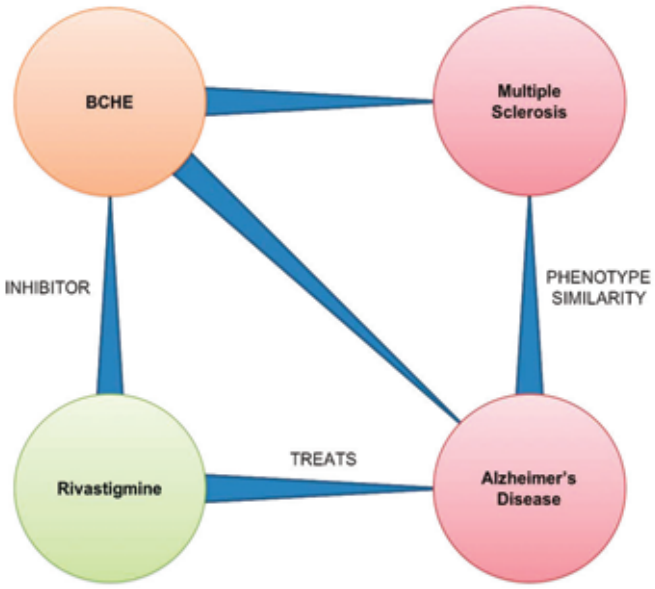

Figure 2: *The potential relationship between Rivastigmine and MS is only revealed through connected data sources*

Document Co-Occurring Indication entities

We've identified other topics of a given type that most often appear in the documents brought back with your query: ***Rivastigmine*** The most frequently occurring topics are listed first in the table below. Document counts under a certain threshold can be scored for relevancy to make triage easier. We recommend only scoring entity-entity intersections under 20 documents as the process is intensive and above 20 articles it's highly likely that there is a meaningful relationship between the two topics.

*Search Term Filtered By: Source: Medline

Scibite Ranking is overlaid. Green > yellow > Red — Overlay Scibite Ranking

1. Alzheimer Disease [1042 docs]	2. Dementia [603 docs]	3. Cognition Disorders [191 docs]
4. Parkinson Disease [188 docs]	5. Psychotic Disorders [149 docs]	6. Depression [78 docs]
7. Vomiting [74 docs]	8. Vascular Dementia [72 docs]	9. Nausea [71 docs]
10. Hallucinations [70 docs]	11. Neurobehavioral Manifestations [70 docs]	12. Memory Disorders [69 docs]
13. Neurodegenerative Diseases [67 docs]	14. neurodegenerative disease [67 docs]	15. Inflammation [60 docs]
16. Delirium [46 docs]	17. Mild Cognitive Impairment [46 docs]	18. Lewy Body Disease [40 docs]
19. Diarrhea [33 docs]	20. Parkinsonian Disorders [32 docs]	21. Schizophrenia [28 docs]
22. Stroke [26 docs]	23. Brain Diseases [23 docs]	24. Amyloid Plaque [23 docs]
25. Dizziness [21 docs]	26. Hypertension [21 docs]	27. Cerebrovascular Disorders [20 docs]
28. Ischemia [20 docs]	29. Bradycardia [19 docs]	30. Aging [18 docs]
31. Tremor [18 docs]	32. Atrophy [17 docs]	33. Multiple Sclerosis [15 docs]
34. Syncope [15 docs]	35. Frontotemporal Dementia [15 docs]	36. Amnesia [14 docs]
37. Brain Injuries [14 docs]	38. Intellectual Disability [14 docs]	39. Anorexia [13 docs]
40. Anorexia Nervosa [13 docs]	41. Diabetes Mellitus [13 docs]	42. Central Nervous System Diseases [12 docs]
43. Brain Ischemia [12 docs]	44. Down Syndrome [12 docs]	45. Neuroblastoma [12 docs]
46. Pain [12 docs]	47. Poisoning [12 docs]	48. Short-Term Memory [11 docs]
49. Movement Disorders [11 docs]	50. Neurotoxicity Syndromes [11 docs]	

First Previous 1 2 3 4 5 Next Last

Figure 3: *A search for Rivastigmine in the biomedical literature using DOCStore[53] ranks Multiple Sclerosis at number 33 in the list of most frequently co-occurring disease indications.*

51 https://www.scibite.com/products/termite/

52 A white paper entitled "Biological knowledge networks and phenotype triangulation" describing the calculation and application of phenotypic similarity measures to semantically enriched data can be obtained from SciBite on request.

53 http://www.scibite.com/products/docstore/

Without an integrated approach which leveraged both structured data sources and unstructured biomedical literature, the identification of MS as a potential repositioning opportunity would have taken significant manual effort and could have easily remained hidden in the ever-increasing volume of biomedical literature.

For validation, a DOCstore sentence co-occurrence search enables us to efficiently focus in on publications mention Rivastigmine in the context of MS. We can see that trials have indeed been conducted into the use of Rivastigmine in the treatment of cognitive processing in patients with MS[54] (Figure 4).

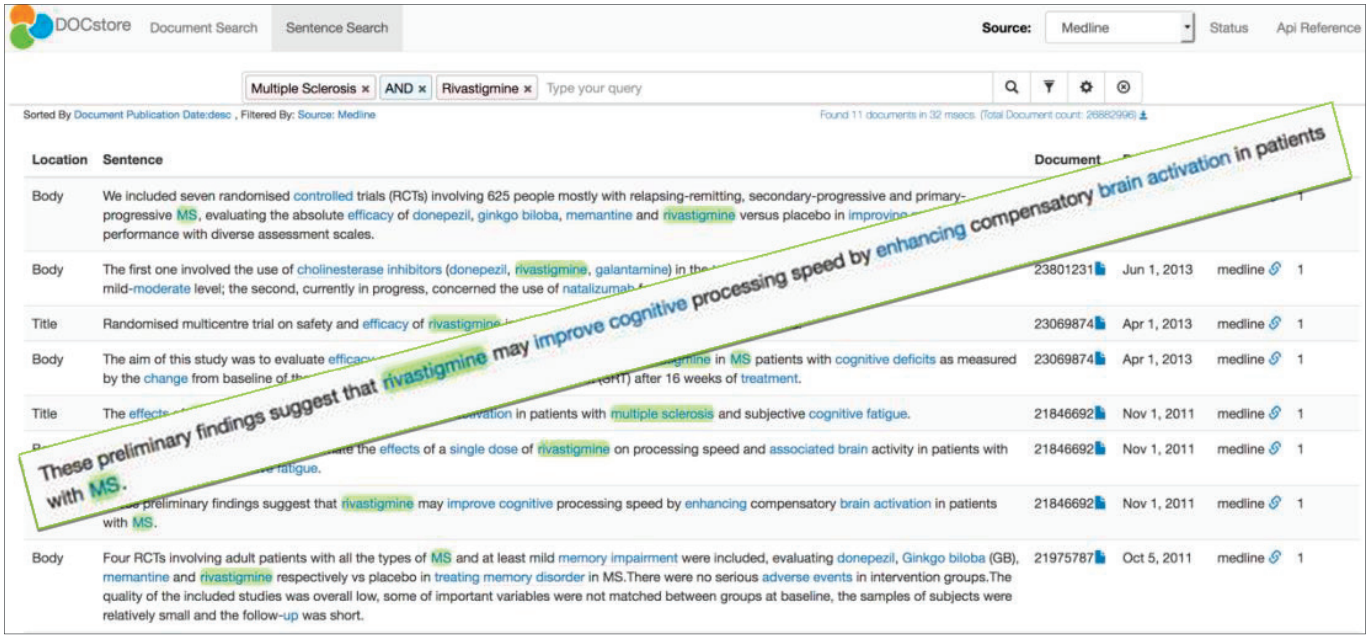

Figure 4: *DOCStore sentence co-ocurrence search to highlight studies around Rivastigmine in the treatment of MS symptoms*

Related applications of semantic analytics

The inclusion of contextualised biomedical literature as part of a holistic approach to Drug Repositioning results in a wealth of context-rich knowledge that can be applied to other areas of Drug Development. For example, the inclusion of safety information in SciBite's semantic network can not only be used to identify diseases with similar side effect profiles, but can also be used as the foundation for an effective Pharmacovigilance surveillance[55]. Better connected data can also identify new opportunities for an existing drug, such as the implications of a newly published drug delivery technology or of changes in regulations in different geographic markets.

Many of the top 10 drugs in the US help as few as 1:25 of patients prescribed them[56], in no small part due to the bias towards white Western participants in clinical trials. Precision medicine is based on the ability to stratify diseases into molecular subtypes and treat patients drugs targeted to their specific disease subtype will improve specificity and efficacy[57]. Rather than treating all patients of a diseases as a homogenous group, the ability to identify and distinguish individual molecular subtypes provides the granularity needed to use semantic networks to identify potential treatments for specific subsets of patients. The increasing availability of detailed patient-level information will provide an opportunity to identify and tailor treatments for an individual[58].

54 Huolman, S. et al (2011). The effects of Rivastigmine on processing speed and brain activation in patients with multiple sclerosis and subjective cognitive fatigue.

55 For example, see http://www.scibite.com/ library-items/semantic-analytics-an-integrated-approach-for-pharmacovigilance-teams-to-achieve-total-awareness-2/

56 Schork N.J. (2015). Personalized medicine: Time for one-person trials. Nature 520:609-611.

57 Li Y.Y., An J. and Jones S.J.M. (2011). A computational approach to finding novel targets for existing drugs. PLoS Computational Biology 7:e1002139.

58 Hodos R.A., Kidd B.A., Shameer J., Readhead B.P. and Dudley J.T. (2016). In silico methods for drug repurposing and pharmacology. WIREs Syst Biol Med. 8(3):186-210.

SciBite
an ELSEVIER company

Press releases and other sources of news are also valuable inputs into a systematic Drug Repositioning strategy. For example, if a news article links a competitor's drug to a specific target and indication, then internally developed drugs that have the same target but are being used to treat a different indication are potential candidates for the competitor's indication.

Semantic networks containing investigator information can help identify relevant expertise, foster collaboration both internally and externally and identify potential external partners. Similarly, greater awareness of the therapeutic and competitive landscape can be achieved by incorporating news and press release information, providing valuable information to help focus an acquisition strategy.

Summary

Drug Repositioning is an efficient and cost effective alternative to de novo drug development and can accelerate the provision of new, safe treatments to underserved patient communities. However, traditional approaches to Drug Repositioning rely on structured data sources, sometimes supplemented by searches of the literature. Due to the manual, time consuming nature of such searches, they are limited in scope and miss potentially important information.

Semantic Analytics facilitates the rapid identification of drug-target-indication relationships from a wide range of heterogeneous and cross-disciplinary sources. The SciBite Platform makes Semantic Analytics accessible to Pharmaceutical companies, expedites the identification and prioritisation of all possible repositioning options associated with one or more drugs of interest and ensures decisions are based on all the available evidence.

SciBite
an ELSEVIER company

SciBite's data-first, semantic analytics software is for those who want to innovate and get more from their data. At SciBite we believe data fuels discovery and we are leading the way with our pioneering infrastructure that combines the latest in machine learning with an ontology-led approach to unlock the value of scientific content. Supporting the world's leading scientific organisations with use-cases from discovery through to development, SciBite's suite of fast, flexible, deployable API technologies empower our customers, making it a critical component in scientific, data-led strategies. Contact us to find out how we can help you get more from your data.

To learn how SciBite can unlock the value of your data, speak to one of our experts today or email us at contact@scibite.com

Head Office:
SciBite Limited
BioData Innovation Centre
Wellcome Genome Campus
Hinxton, Cambridge CB10 1DR
United Kingdom

www.scibite.com
contact@scibite.com
LinkedIn: SciBite
Twitter: @SciBite
+44 (0)1223 786 129