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

Editors-in-Chief
Raymond Baker – formerly University of Southampton, UK and Merck Sharp & Dohme, UK
Eliot Ohlstein – GlaxoSmithKline, USA

Drug repurposing

Literature analysis for systematic drug repurposing: a case study from Biovista

Eftychia Lekka, Spyros N. Deftereos, Aris Persidis, Andreas Persidis, Christos Andronis*

Biovista Inc., 2421 Ivy Road, Charlottesville, VA 22903, USA

Drug Repositioning (DR) has recently emerged as a complementary approach to classic drug discovery aiming at challenging the limited productivity issues associated with the traditional drug discovery route. By seeking novel links between existing drugs and new indications, data mining of various sources constitutes a powerful tool for systematic DR. This review focuses on primary literature as one of the data sources and on Literature-Based Discovery (LBD) strategies for DR, presenting a relevant case study for the treatment of Multiple Sclerosis (MS).

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

Introduction

De novo drug development is laborious (15–20 years) and costly (500 million to 2 billion US dollars) [1–3]. The pharmaceutical industry is constantly seeking for ways to diminish costs without sacrificing its productivity. By contrast, the number of New Chemical Entities (NCEs) approved by the Food and Drug Administration (FDA) each year remains constant at 20–30 compounds [4]. On top of the mounting R&D costs, pharmaceutical companies are currently under the strain of other issues, including loss of revenue due to patent expirations, increasingly cost-constrained healthcare systems and regulatory hurdles [5]. All these factors have forced the bio-pharmaceutical industry toward complementary productivity strategies, including in-licensing, merger-and-acquisition and, finally, Drug Repositioning (DR, also called indication expansion, indication switching, drug repurposing or drug reprofiling) [6].

As its name implies, DR focuses on identifying new indications for existing drugs and has been employed since the first steps of the pharmaceutical industry, mainly through clinical observation and serendipity, with well-known examples being *sildenafil* and *thalidomide*. DR candidates share some common features; have either exhibited a prior lack of efficacy in clinical trials (without safety issues) or stalled in development for commercial reasons and, in case of approved drugs, might have been subjected to patent expiry, or seen as a market strategy in other geographical areas [7].

One would think that the witnessed explosion in high-throughput data, particularly in the genomics field, would lead to a burst in discovery of new drugs. However, the number of '*druggable targets*' [8] has not increased proportionately as a result of the sequencing of the human genome. By utilizing existing drugs for new indications, DR helps reshuffle the drug target space.

DR is an appealing approach also from the patient's perspective, because the repositioned drug has usually a well-established safety profile. This confers not only significant time gains but also reduction in attrition, because safety concerns are by definition fewer compared to those involving a new drug [9].

Although DR strategies can make use of a variety of data sources and data mining approaches, here we review

*Corresponding author.: C. Andronis (candronis@biovista.com)

1740-6773/$ © 2011 Elsevier Ltd. All rights reserved. DOI: 10.1016/j.ddstr.2011.06.005

systematic literature-based DR strategies and describe the literature part of the DR approach followed by Biovista, including a prominent example for a repositioned drug in Multiple Sclerosis (MS).

Technology background

Literature mining and its application in the biomedical setting

Part of the process of modern biomedical research involves the analysis of large volumes of experimental data to identify new patterns or correlations, ultimately leading to novel hypotheses. Data mining methodologies have been extensively used to infer knowledge from high-throughput genomic, chemical, metabolic and proteomic data [10].

The scientific literature can also be thought as another form of a data repository capturing biomedical facts and their interrelations. With the volume of biomedical literature constantly increasing, it is becoming difficult to locate, retrieve and manage the reported information without literature mining. Literature mining uses text mining tools to automatically extract facts and possible relations, thus enabling users to systematically generate and substantiate appropriate hypotheses [11]. Systematic literature analysis offers a methodical understanding of a disease's underlying mechanisms and helps monitoring areas where scientific advances are likely to create new therapeutic opportunities [12].

New knowledge can be gained by mining the scientific literature for direct or indirect relationships. The process of generating novel hypotheses by linking seemingly unrelated facts or indirect connections is commonly referred to as *Literature-Based Discovery* (*LBD*) [13]. LBD is based on the assumption that two 'islands of knowledge' A and C might be related to each other if they share a link (in the literature, or in other repositories of empirical data) with an intermediate concept B (Fig. 1). The term '*ABC model*' has been coined to describe this mode of scientific discovery through inference. The initial discoveries made by Swanson [14–16] were based on a manual scanning of the literature; however, the immense expansion of the scientific literature has necessitated the use of Information Extraction (IE) technologies to extract terms and concepts of interest from free-text.

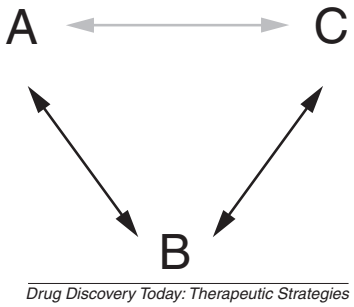


Figure 1. Swanson's ABC model of discovery. If concepts A and B are reported to be related to one set of publications and concepts B and C are reported to be related to another set, then A and C might be indirectly related to each other.

IE comprises a first step of Entity Recognition (ER) to identify biomedical terms of interest in free-text (e.g. a Medline abstract) and subsequently link these terms to each other using either *co-occurrence* or more sophisticated *Natural Language Processing* (*NLP*)-based methods [17,18]. Co-occurrence, which associates terms with each other when they appear in the same text, has been extensively used in IE systems in the context of literature-based discoveries [17,19,20]. Although prone to generating false positives and unable to provide information regarding the nature of the relationship, co-occurrence can be well applied as an exploratory component, as it can provide associations of almost any type and does not distinguish between direct and indirect relationships [17,21]. Another key factor for the prevalence of co-occurrence is its robustness in the context of large-scale IE systems.

Drug discovery by systematic literature analysis

Following Swanson's initial LBD discoveries regarding dietary fish oil as a potential therapeutic for Raynaud's disease [14], numerous variations of the original ABC model have appeared in the literature [22–25]. Notable examples include the suggestion of *thalidomide* for the treatment of chronic hepatitis C [26] and the 'Litlinker' approach [27] which identified three new connections: endocannabinoids with Alzheimer's Disease (AD), AMPA receptors with migraine, and secretin with schizophrenia.

Recent examples using literature mining for systematic DR include, among others, work by Li *et al.* [28], which used mined information from PubMed abstracts and molecular interaction networks to construct drug–protein connectivity maps. Development and application of this computational framework were performed on AD, as a starting example, leading to a new hypothesis that *diltiazem* and *quinidine* could be investigated as candidate drugs for AD treatment. Moreover, Frijters *et al.* [29] constructed an 'ABC'-based literature mining tool, called 'CoPub Discovery', to identify novel connections between drugs, genes and diseases. Using co-occurrence, they created interconnections between gene names and other biomedical concepts, which they extracted from Medline abstracts. Co-citation strength was assessed using a mutual information-based metric and new relations were then validated through a series of case studies and *in vitro* experiments, such as the finding that dephostatin and damnacanthal, a tyrosine phosphatase inhibitor and a tyrosine kinase inhibitor, respectively, are associated with cell proliferation. Recently, von Eichborn *et al.* [30] created a network-based resource of protein–protein and protein–drug interactions, called 'PROMISCUOUS', which they enriched with side-effect and structural information data mined from appropriate databases, aiming at providing a uniform data set

for further analysis and a good start for DR, integrating basic graph theoretical analysis methods.

Drug Repositioning: a Biovista case study for Multiple Sclerosis

Information extraction

Biovista (Charlottesville, VA, USA) utilizes data mining techniques to identify nonobvious relationships among drugs, new indications and potential Adverse Drug Reactions (ADRs). Part of Biovista's discovery workflow is based on Systems Literature Analysis (SLA) [31], which treats the scientific literature as a vast system of interconnections between various biomedical entities. SLA-based discovery consists of integrated IE tools, a database of relations between biomedical entities and inferential algorithms rooted in LBD. Clinical Outcome Search Space™ (COSS) is Biovista's computational platform which incorporates the SLA engine, as well as data analysis and visualization tools (Fig. 2). COSS relies on an extensive proprietary database of context-crossing relations among biomedical entities (diseases, ADRs, drugs, compounds, genes, pathways, among others), which utilizes custom technological solutions and correlational tools to achieve high performance access to the underlying data and their interrelations. The process is strengthened by the integration of cheminformatics tools and related data sources [21,32]. This enables the user not only to generate information regarding the putative Mode of Action (MoA) of drugs in novel indications but also to extract information regarding potential ADRs. COSS enables Biovista to repurpose drugs and drug targets and assess their Benefit/Risk clinical outcome potential over the entire range of medical conditions listed in the various Unified Medical Language System (UMLS)

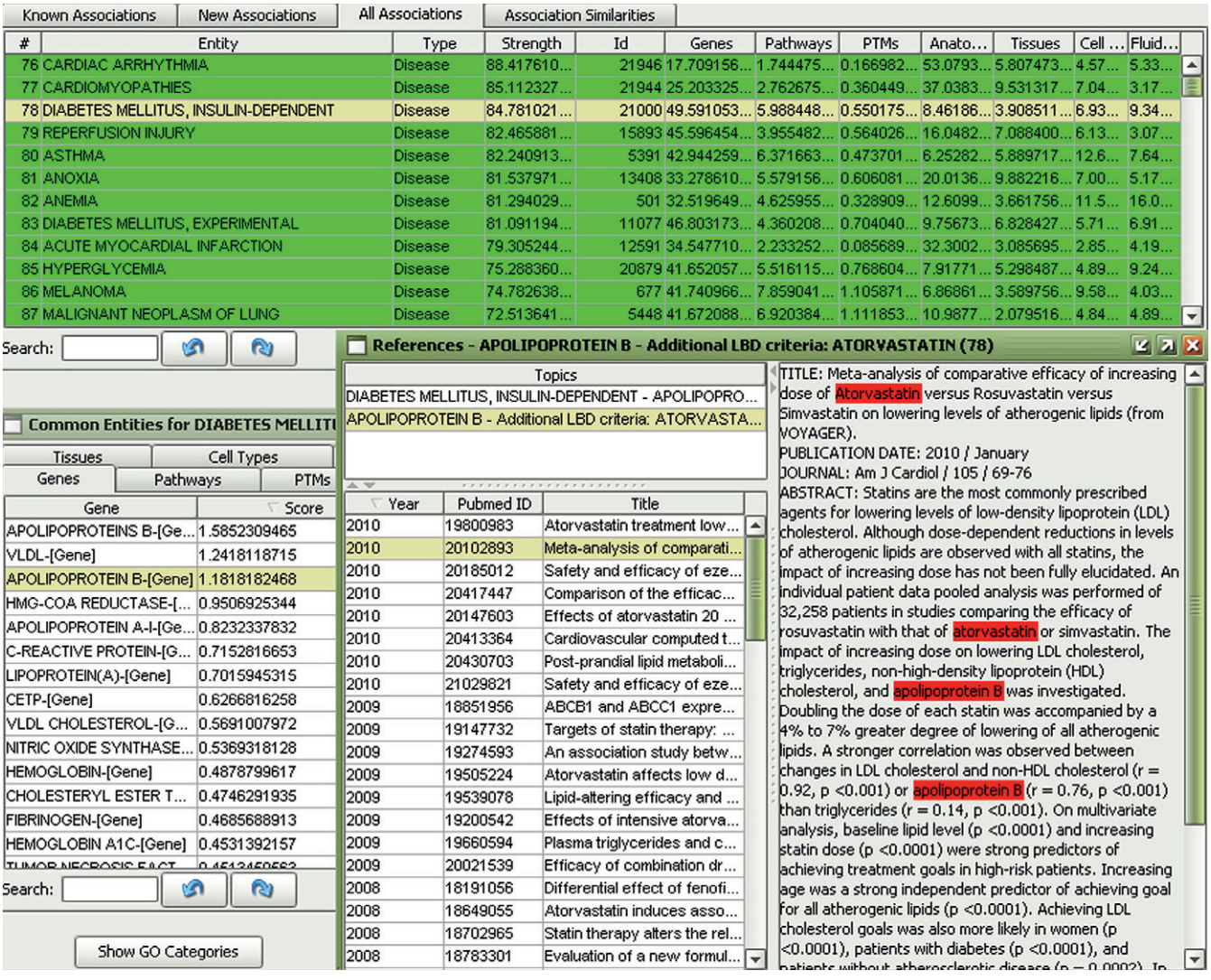


Figure 2. Screenshot of the analysis module of Biovista's COSS platform. The figure displays diseases related to the drug atorvastatin, as an example. Terms relating atorvastatin to the various diseases are shown together with the relevant bibliography.

Knowledge sources. This technology is currently employed for the development of the company's pipeline in areas such as central nervous system (CNS), Cardiovascular Disease, Oncology and Auto-immune diseases.

Resources

COSS incorporates data from literature corpora such as PubMed [33], bioinformatics databases and ontologies.

Biomedical data are retrieved from the following resources:

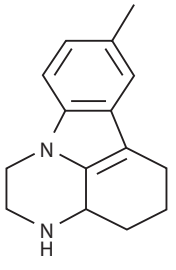
(a) *Genomic databases* include EntrezGene [34], the Universal Protein Resource (UniProt) [35], Protein Data Bank (PDB) [36] and Gene Expression Atlas [37].
(b) *Microarray repositories*, such as Gene Expression Omnibus (GEO) [38] and ArrayExpress [39].
(c) *Pathway databases*, such as KEGG Pathway (Kyoto Encyclopedia of Genes and Genomes) [40], Reactome [41] and the National Cancer Institute (NCI)-Nature Pathway Interaction Database [42].
(d) *Cheminformatics resources*, such as PubChem [43], ChEMBL [44] and ChemSpider [45].
(e) *Drug-to-target databases* include DrugBank [46], PharmGKB [47] and MATADOR (Manually Annotated TArgets and Drugs Online Resource) [48].
(f) *Drug-related-to-disease databases* include the US Food and Drug Administration Drugs@FDA database [49], and others, designed to support FDA's postmarketing safety surveillance, such as Adverse Event Reporting System (AERS) [50], Side Effect Resource (SIDER) [51] and DailyMed [52].

In addition, COSS exploits biomedical ontologies such as the UMLS [53] from the US National Library of Medicine (NLM), and the Medical Subject Headings thesaurus (MeSH) [54], which is a controlled vocabulary for biomedical terms also from the NLM.

Discovery

BVA-201: repositioning overview

Using the concepts laid above for SLA and COSS, Biovista has repositioned Pirlindol (code-named to BVA-201, Fig. 3) to MS.



Drug Discovery Today: Therapeutic Strategies

Figure 3. The chemical structure of Pirlindol.

The drug, which is a reversible inhibitor of Monoamine Oxidase A (RIMA), was initially approved in Russia and the European Union for the chronic treatment of depression, affective and psychotic disorders with a depressive component, and of anxiety disorders. Biovista has filed patents for BVA-201 for the treatment of MS [55].

Multiple Sclerosis: disease overview and current therapeutic strategies

MS is an inflammatory-mediated demyelinating disorder of the CNS of unknown etiology [56], which affects around 2.5 million people worldwide. As the disease progresses, neurological decline is evident and this is attributed to mechanisms independent of an adaptive immune response which closely resemble neurodegeneration. Northern Europe, the northern United States, southern Australia and New Zealand have the highest prevalence, with more than 30 affected cases per 100,000 people and an increased incidence in women [57]. Most MS drugs are currently targeting the inflammatory component of the disease while not having any effect on neurodegeneration. Although drugs in this class provide a significant reduction in the rate of relapses (in the order of 30–67%), they are associated with mild-to-often fatal adverse events, such as in the case of Natalizumab which induces Progressive Multifocal Leukoencephalopathy (PML), a rare, but life-threatening disease [58].

BVA-201, as evident from clinical trials and postmarketing surveillance, has a very good adverse event profile in humans, exhibiting only mild and transient adverse events including dry mouth, increased sweating, hand tremor, tachycardia, nausea and dizziness. In clinical practice, the drug does not seem to interfere with the gastrointestinal tract function, heart rate, blood pressure, ECG, water–electrolyte balance [59]. Furthermore, the drug can be used by patients who cannot tolerate tricyclic antidepressants, including patients with prostate hypertrophy and glaucoma [59], and patients with cardiovascular co-morbidities, in which it has no interactions with commonly prescribed drugs [60].

Experimental validation of BVA-201 role in an animal model of MS

Apart from acting as a MAO reversible inhibitor, the MoA of BVA-201 involves the reduction of oxidative stress and inhibition of lipid peroxidation [61,62], a process involved both in relapsing–remitting and in the progressive forms of MS. This behavior is not seen in other MAO inhibitors [61], implying that the anti-oxidant effect of BVA-201 is specific and not a generic characteristic of drugs within its class. In *in vivo* experiments performed in the MOG (Myelin Oligodendrocyte Glycoprotein)-induced Experimental Allergic Encephalomyelitis (EAE) murine model of MS, BVA-201, at 30 mg/kg (a pharmaceutically relevant dose) induced a sizeable improvement in disease progression, as represented

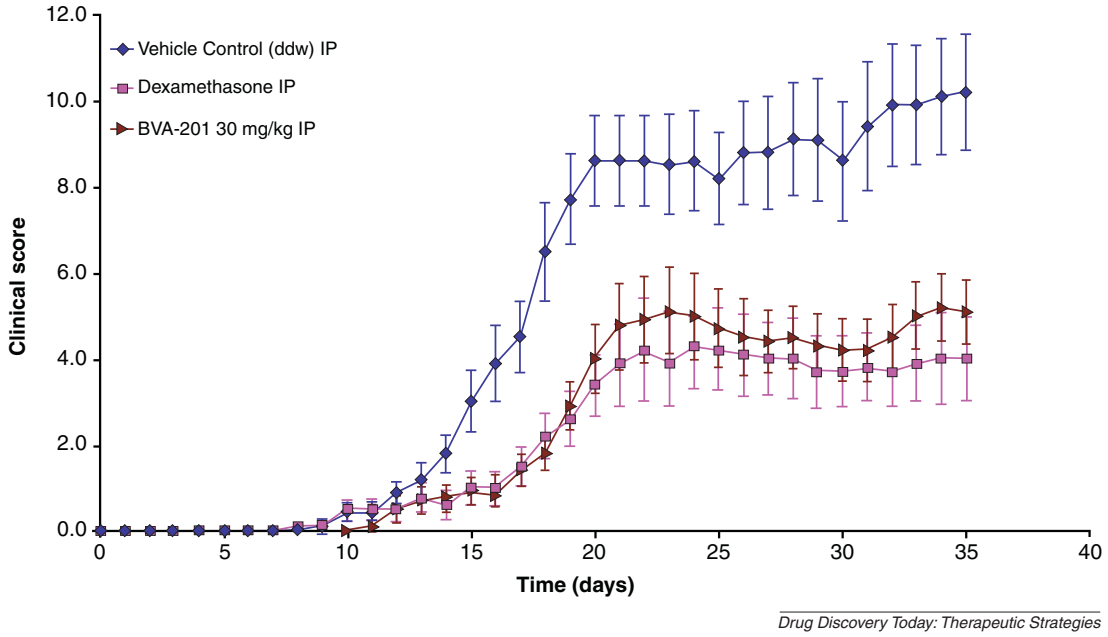


Figure 4. Disease progression in an experimental model of Multiple Sclerosis. BVA-201 (30 mg/kg) induced a statistically significant reduction of EAE severity starting on day 11 and maintained throughout the rest of the experiment (day 35). On day 35, the average score among experimental groups was 5.10 $\pm$ 2.3 in the case of the 30 mg/kg group, 4.0 $\pm$ 3.1 in the positive control group, while the mean score of vehicles was 10.2 $\pm$ 4.28. The difference between the BVA-201 30 mg/kg group and the vehicle group was highly statistically significant ($P < 0.01$), while the difference between the BVA-201 30 mg/kg group and the positive control group was not statistically significant ($P = 0.38$). Lower clinical scores correspond to less severe disease.

by EAE severity clinical scores (Fig. 4). Moreover, histological data showed that the drug's MoA is not anti-inflammatory, but rather, BVA-201 protects neuronal cells and myelin in a direct manner. These observations lead to the validation of the hypothesis involving repositioning BVA-201 for MS.

Biovista's adverse event prediction

Along with DR, Biovista utilizes its COSS platform for predicting ADRs. Prediction of the ADRs allows improved clinical trial design and optimal patient stratification during clinical development, and can translate into better understanding of a compound's risk profile, ultimately leading to accelerated development and significant savings. Drug-related ADR information is incorporated into COSS from databases such as AERS and SIDER and, more importantly, from published data.

Conclusion

DR strategies have been increasingly fruitful in recent years; in 2009 alone, more than 30% of the 51 new medicines launched in the market involved new indications, new formulations and new combinations of previously marketed drugs [63]. Being a cost and time saving route, DR is expected to play an increasingly major role in the pipelines of biopharmaceutical companies.

At the same time, the notion of linking an existing drug and, hence, a known drug target to a new indication is a very promising approach to uncover hidden connections between previously unconnected signaling pathways, in turn, leading to 'novel biology'. Overlooked pathways and pharmacological mechanisms leading to new indications are uncovered by the drug's ability to bind either to a known target (on-target effect) or a 'novel' target for which the drug was not originally designed for (off-target effect). Either way, for every piece of information added to the corpus of existing scientific knowledge, DR is the implement which combines prior and acquired knowledge, thus creating a large input of data that could readily lead to innovation.

The recent success stories of DR have spawned the creation of several discovery programs attempting to systematize the application of existing drugs to new indications. Among the various platforms suggested in the literature, *in silico* DR based on literature mining, seems to be a promising strategy for such a goal. The present review has presented a compelling DR-driven hypothesis for the use of an existing drug for a new indication, subsequently validated in an *in vivo* model of the disease.

References

1 Chong, C.R. and Sullivan, D.J., Jr (2007) New uses for old drugs. *Nature* 448, 645–646

2 Adams, C.P. and Brantner, V.V. (2006) Estimating the cost of new drug development: is it really 802 million dollars? *Health Aff. (Millwood)* 25, 420–428

3 DiMasi, J.A. *et al.* (2003) The price of innovation: new estimates of drug development costs. *J. Health Econ.* 22, 151–185
4 Cohen, F.J. (2005) Macro trends in pharmaceutical innovation. *Nat. Rev. Drug Discov.* 4, 78–84
5 Paul, S.M. *et al.* (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9, 203–214
6 Reilly, T. and Yost, N. (2006) Drug Repositioning. *Res. Dev. Strat. Drug Discov.* 10–12. Report at http://www.touchbriefings.com
7 Sleigh, S.H. and Barton, C.L. (2010) Repurposing strategies for therapeutics. *Pharm. Med.* 24, 151–159
8 Hopkins, A.L. and Groom, C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730
9 Persidis, A. (2011) Drug Repositioning: a union of patient interests, pipeline development and innovation. *Drug Discovery & Development – March 29*
10 Yang, Y. *et al.* (2009) Target discovery from data mining approaches. *Drug Discov. Today* 14, 147–154
11 Spasic, I. *et al.* (2005) Text mining and ontologies in biomedicine: making sense of raw text. *Brief. Bioinform.* 6, 239–251
12 Agarwal, P. and Searls, D.B. (2009) Can literature analysis identify innovation drivers in drug discovery? *Nat. Rev. Drug Discov.* 8, 865–878
13 Weeber, M. *et al.* (2005) Online tools to support literature-based discovery in the life sciences. *Brief. Bioinform.* 6, 277–286
14 Swanson, D.R. (1986) Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect. Biol. Med.* 30, 7–18
15 Swanson, D.R. (1988) Migraine and magnesium: eleven neglected connections. *Perspect. Biol. Med.* 31, 526–557
16 Swanson, D.R. (1990) Medical literature as a potential source of new knowledge. *Bull. Med. Libr. Assoc.* 78, 29–37
17 Jensen, L.J. *et al.* (2006) Literature mining for the biologist: from information retrieval to biological discovery. *Nat. Rev. Genet.* 7, 119–129
18 Saric, J. *et al.* (2006) Extraction of regulatory gene/protein networks from Medline. *Bioinformatics* 22, 645–650
19 Hale, R. (2005) Text mining: getting more value from literature resources. *Drug Discov. Today* 10, 377–379
20 Jenssen, T.K. *et al.* (2001) A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.* 28, 21–28
21 Andronis, C. *et al.* (2011) Literature mining, ontologies and information visualization for drug repurposing. *Brief. Bioinform.* 12, 357–368
22 Gordon, M.D. and Lindsay, R.K. (1996) Toward discovery support systems: a replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil. *J. Am. Soc. Inf. Sci.* 47, 116–128
23 Hristovski, D. *et al.* (2001) Supporting discovery in medicine by association rule mining in Medline and UMLS. *Proceedings of Med. Info Conference (Vol. 10)*, London, England, 2–5 September. pp. 1344–1348
24 Srinivasan, P. (2001) MeSHmap: a text mining tool for MEDLINE. *Proc. AMIA Symp.*pp. 642–646
25 Weeber, M. *et al.* (2001) Using concepts in literature-based discovery: simulating Swanson's Raynaud-fish oil and migraine-magnesium discoveries. *J. Am. Soc. Inf. Sci.* 52, 548–557
26 Weeber, M. *et al.* (2003) Generating hypotheses by discovering implicit associations in the literature: a case report of a search for new potential therapeutic uses for thalidomide. *Am. Med. Inform. Assoc.* 10, 252–259
27 Yetisgen-Yildiz, M. and Pratt, W. (2006) Using statistical and knowledge-based approaches for literature-based discovery. *J. Biomed. Inform.* 39, 600–611
28 Li, J. *et al.* (2009) Building disease-specific drug-protein connectivity maps from molecular interaction networks and PubMed abstracts. *PLoS Comput. Biol.* 5, e1000450
29 Frijters, R. *et al.* (2010) Literature mining for the discovery of hidden connections between drugs, genes and diseases. *PLoS Comput. Biol.* 6, e1000943
30 von Eichborn, J. *et al.* (2011) PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res.* 39 (Database issue), D1060–D1066
31 Persidis, A. *et al.* (2004) Systems literature analysis. *Pharmacogenomics* 5, 943–947
32 Deftereos, S.N. *et al.* (2011) Drug repurposing and adverse event prediction using high-throughput literature analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* Feb 16 (Epub ahead of print)
33 PubMed available at: http://www.ncbi.nlm.nih.gov/pubmed
34 Entrez Gene available at: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene
35 UniProt Consortium, (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* 38 (Database issue), D142–D148
36 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
37 Gene Expression Atlas available at: http://www.ebi.ac.uk/gxa/
38 Gene Expression Omnibus (GEO) available at: http://www.ncbi.nlm.nih.gov/geo/
39 Array Express available at: http://www.ebi.ac.uk/arrayexpress/
40 Kanehisa, M. *et al.* (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34 (Database issue), D354–D357
41 Matthews, L. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 37 (Database issue), D619–D622. Available at: http://www.reactome.org/ReactomeGWT/entrypoint.html
42 Pathway Interaction Database available at: http://pid.nci.nih.gov
43 PubChem available at: http://pubchem.ncbi.nlm.nih.gov/
44 ChEMBL available at: http://www.ebi.ac.uk/chembldb/
45 ChemSpider available at: http://www.chemspider.com/
46 DrugBank available at: http://www.drugbank.ca
47 Hewett, M. *et al.* (2002) PharmGKB: the Pharmacogenetics Knowledge Base. *Nucleic Acids Res.* 30, 163–165. Available at: http://www.pharmgkb.org/
48 Günther, S. *et al.* (2008) SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.* 36 (Database issue), D919–D922. Available at: http://matador.embl.de/
49 Drugs@FDA available at: http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm
50 Adverse Event Reporting System (AERS) available at: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm
51 SIDER available at: http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm
52 DailyMed available at: http://dailymed.nlm.nih.gov/
53 UMLS available at: http://www.nlm.nih.gov/research/umls/
54 MeSH available at: http://www.nlm.nih.gov/mesh/meshhome.html
55 Deftereos, S.N. *et al.* [Biovista] (2010) Methods for treating Multiple Sclerosis using tetracyclic pyrazinoindoles. *WIPO Patent Application WO/2010/068867*
56 Noseworthy, J.H. *et al.* (2000) Multiple sclerosis. *N. Engl. J. Med.* 343, 938–952
57 Debouverie, M. (2009) Gender as a prognostic factor and its impact on the incidence of multiple sclerosis in Lorraine, France. *J. Neurol. Sci.* 286, 14–17
58 Tysabri Package Insert. Available at: http://www.tysabri.com/en_US/tysb/site/pdfs/TYSABRI-pi.pdf
59 Mozorov, P.V. (2002) Pirazidol – effective domestic anti-depressant. *Consilium Medicum* 4, N6
60 Ivanov, S.V. *et al.* (2003) Pyrazidol in the treatment of depression in patients with ischemic heart disease. *Ter. Arkh.* 75, 38–42
61 Boland, A. *et al.* (2003) Pre- and post-treatment with pirlindole and dehydropirlindole protects cultured brain cells against nitric oxide-induced death. *Eur. J. Pharmacol.* 466, 21–30
62 Boland, A. *et al.* (2002) Pirlindole and dehydropirlindole protect rat cultured neuronal cells against oxidative stress-induced cell death through a mechanism unrelated to MAO-A inhibition. *Br. J. Pharmacol.* 135, 713–720
63 Graul, A.I. *et al.* (2010) The year's new drugs & biologics – 2009. *Drug News Perspect.* 23, 7