ELSEVIER

Network analysis has diverse roles in drug discovery

Samiul Hasan, Bhushan K. Bonde, Natalie S. Buchan and Matthew D. Hall

GlaxoSmithKline, Computational Biology, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK

Computational biologists use network analysis to uncover relationships between various data types of interest for drug discovery. For example, signalling and metabolic pathways are commonly used to understand disease states and drug mechanisms. However, several other flavours of network analysis techniques are also applicable in a drug discovery context. Recent advances include networks that encompass relationships between diseases, molecular mechanisms and gene targets. Even social networks that mirror interactions within the scientific community are helping to foster collaborations and novel research. We review how these different types of network analysis approaches facilitate drug discovery and their associated challenges.

Introduction

Drug discovery is a long and complex process requiring multidisciplinary approaches to develop safe and efficacious medicines. The amount and sheer diversity of data generated and used, as a discovery project moves through the different phases of the pipeline, is enormous. The analysis and tracking of this data for pipeline progression is a massive challenge. However, it is also pertinent to ask whether taking a more overarching view of this wealth of data might uncover underlying patterns that could positively influence the drug discovery process.

Network analysis, a branch of computational biology, offers approaches for taking different kinds of entities ('nodes'), modelling the relationships ('edges') between these and generating hypotheses from the resulting network. Because drug discovery involves integrative analysis of many different types of data, it is a natural fit for this type of approach. Major nodes of interest for drug discovery can include drugs and their targets, diseases, molecular pathways, published observations, clinical and/or phenotypic outcomes and even people such as key opinion leaders (KOLs) (Fig. 1). In this article we review recent advances in trying to visualise and analyse these using network methodologies and explain how the analysis can facilitate drug discovery.

Corresponding author: Hasan, S. (samiul.x.hasan@gsk.com)

Biological networks

Biological networks are commonly used frameworks to map metabolic and signalling pathway knowledge. On their own, they provide a useful way to share and improve understanding about well-described biological processes. However, with advances in network analysis tools and the advent of systems biology approaches, there is strong interest from a mathematical perspective in extending these static maps to determine key properties of pathways. This has the capacity to refine biological knowledge and hypotheses and provide better decision making for therapeutic interventions.

Genome-scale metabolic networks

The past decade has seen the rise of genome-scale metabolic (GSM) networks (Fig. 1) [1]. GSMs are designed to model metabolism at a cellular level. A complementary computational modelling approach called 'Flux Balance Analysis' (FBA) is used to study GSMs. FBA uses stoichiometric constraints and rates of extracellular metabolite uptake and production as input. In FBA, the GSM network is modelled as a system in a pseudo-steady state. This means that the growth rate of a cell is assumed to be constant. One advantage of FBA is that it does not require prior knowledge of enzyme kinetics and/or concentrations of intracellular metabolites. These models provide a framework for computationally interpreting rates of intracellular reactions. This has been applied for studying organisms of various complexities ranging from

1359-6446/06/$ - see front matter http://dx.doi.org/10.1016/j.drudis.2012.05.006

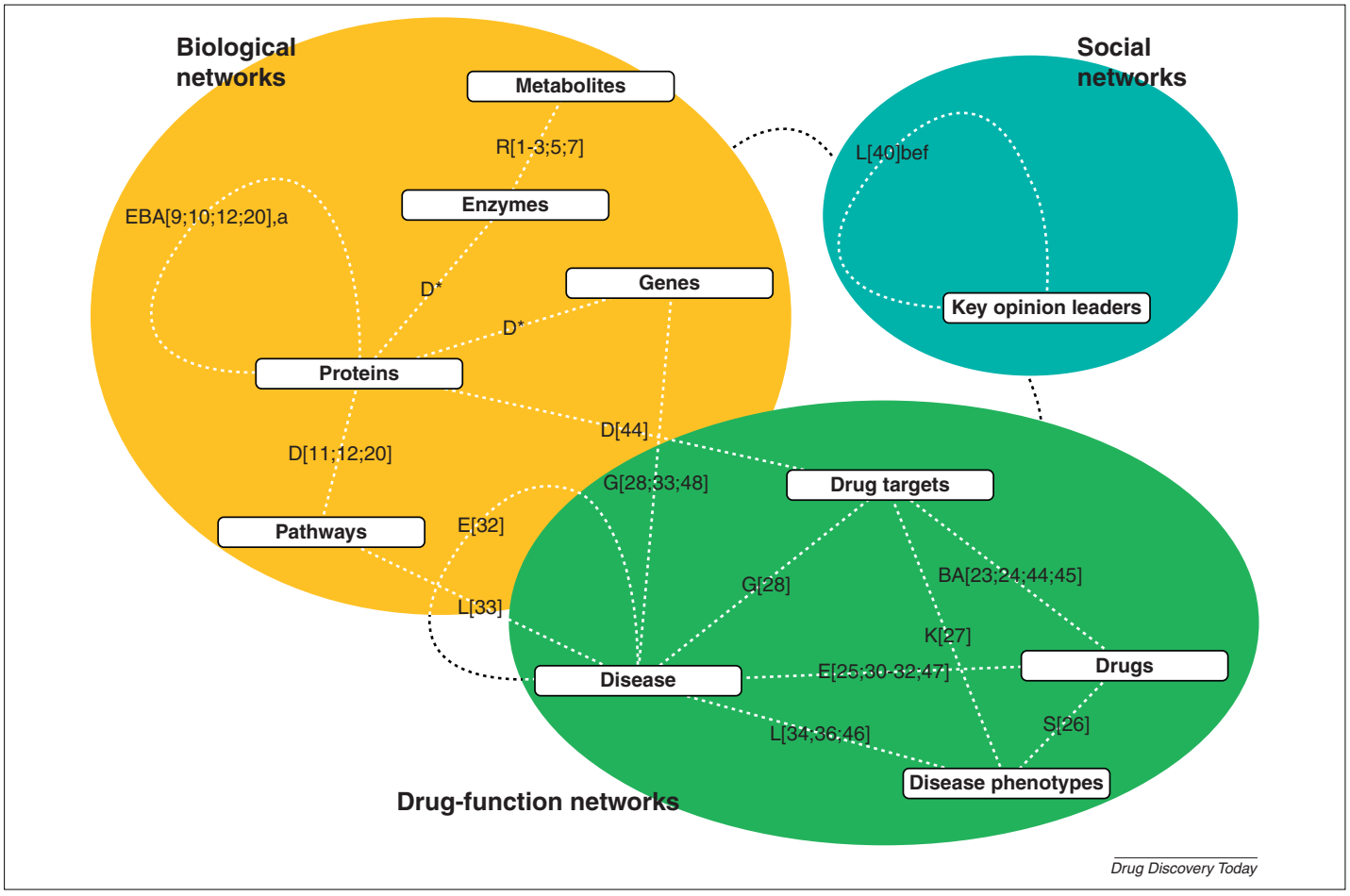

FIGURE 1

Essential entities ('nodes') of interest for drug discovery and approaches to determine their associations. Relationship ('edges') are as follows: A – molecular activity; B – physical binding; D – database repository; E – gene expression; G – genetic association; K – knockout phenotype; L – literature-based including ontologies and co-citations; R – metabolic reaction; S – side effect.

microbial pathogens [2] to higher organisms, including mouse and human [3]. This technique is also being applied in the biopharmaceuticals field for optimising monoclonal antibody production in Chinese hamster ovary (CHO) cell lines [4].

Computational tools such as differential producibility analysis (DPA) [5] have been developed to predict mechanisms of virulence and persistence in *Mycobacterium tuberculosis* and other important pathogens. Such methods help to integrate transcriptional data with the metabolic network. DPA uses the FBA approach to identify sets of genes, that when knocked out, affects the ability of the cell to produce individual metabolites. This is followed by statistical analysis of the expression data to identify and rank key metabolic products. Recently Li *et al.* [6] used the FBA method for identifying new drug target sites in the host pathogen network. Folger *et al.* [7] used a human GSM to model cancer metabolism. This model was used to predict new cytostatic drug targets that inhibit cancerous cell growth. A total of 40% of the proteins, identified by this approach, were known anticancer drug targets thus demonstrating a potential application for drug discovery.

The biggest challenge in developing GSMs is that building such models requires integration of large data sets from various -omics platforms, such as microarrays and proteomics. This is usually a time-consuming process and requires extensive cross-validation. It is also difficult to know the exact metabolites, which are produced by a cell, in various tissues and disease states. In practice, building and modelling GSMs requires a dedicated resource, which can be difficult to obtain in a drug industry setting. Another limitation is that most of the GSMs, which have been developed until now, have been based on FBA-determined criteria. Many research groups also disagree about whether the underlying pseudo-steady state assumption is valid. They argue that this makes it difficult to infer exact reaction kinetics, which would better reflect *in vivo* activity [8]. Also, despite the extensive size and detailed annotation of current GSMs, their application only models relationships between proteins and reactions. To better depict reality, they should be extended to include protein–protein interactions (PPIs) and the influence of signalling molecules. Overall, the development of a good metabolite resource would greatly benefit studies in this area.

Protein interaction and signalling networks

PPIs are another type of biological relationship that can be modelled using network analysis. PPIs describe various binding actions between proteins and include post-translational modifications and dimerization events. These are referred to as 'direct interactions'. By contrast, protein signalling networks (PSN) describe the

impact of transcription factors and other signalling molecules on downstream events including changes in gene expression and post-translational states of proteins [9]. These molecules do not need to physically bind to their target genes to alter their expression; hence, they impart 'indirect interactions'. Both PPI and PSN interactions are described in published literature (Fig. 1) and are derived from a variety of techniques including yeast two-hybrid and other omics approaches. Literature mining is used to systematically search for these individual interactions, with the results captured in databases. Examples of such databases include IntAct [10], Wikipathways [11], Reactome [12] and Ingenuity (Ingenuity Systems: http://www.ingenuity.com). Recently, host–pathogen PPI networks have also been developed [13]. These networks provide significant insights into the host–pathogen interactions during infection state and provide novel therapeutic intervention points for infectious disease. The use of standardisation across databases to enable integration and sharing of data is crucial, and efforts such as BioPAX are essential [14].

Context-specific queries of these interaction databases are used to generate biological hypotheses that will need to be tested experimentally. Typical examples include querying databases with lists of genes, which were found to be differentially regulated in -omics experiments. Computational algorithms help to discover statistically significant sub-networks that are enriched with the genes of interest. In a recent example, Nibbe *et al.* used PPI knowledge and mRNA expression analysis to discover functional sub-networks that may be associated with colorectal cancer progression [15]. This was done by applying a novel computational algorithm to identify sub-networks of a PPI network whose genes were differentially expressed in the disease state. Furthermore, it was recently demonstrated that differentially expressed sub-network markers were more accurate than single gene markers in predicting metastasis in breast cancer [16]. PPI network analysis, therefore, provides insights into the underlying biology and can help to understand disease mechanisms and identify candidate biomarkers. They can also help to predict which genes regulate a drug target or are affected downstream of the target. In analogy to applying FBA to metabolic networks, mathematical modelling of PSNs is also an active area of research. These help to predict the function of PSNs and may be applied to predict PSN responses to specific stimuli [17].

The major problem with PPI and PSNs is that the interactions themselves are determined under defined experimental conditions. This makes it difficult to extrapolate the networks to other environmental and cellular conditions. The other issue with PPI data is missing information. However, these efforts are important steps as they provide a glimpse of the cellular system in a given state. If we can understand these properly, we can begin to study how they change under different conditions. A recent review [18] highlights efforts to study the dynamics of biological networks and emphasises that these studies may become the de-facto network modelling approach in the future.

At the time of writing there are currently approximately 50,000 binary human interactions in the IntAct database [10], which represents just under 10% of the estimated human interactome [19]. Protein modification events are also rarely captured in public databases; however, the Reactome pathway database [20] has begun to address this. Individual interactions must also be manually checked to discard erroneous literature-mining results. Furthermore, binding databases often store negative findings (*e.g.* 'lack of binding') as an interaction and these need to be carefully filtered out.

Drug function networks

Drugs are designed to block the onset or progression of disease but inevitably illicit several other biological responses in patients. Some of these changes may contribute to undesirable side effects including toxicity whereas some may be benign in nature. Knowledge of the latter may help to identify new indications for an existing drug, a term commonly referred to as 'drug repositioning' [21]. In both cases, a single drug is causally linked at the molecular level to several possible phenotypic outcomes. This concept forms the basis of drug function networks. There are two main areas where drug function networks can benefit drug discovery: polypharmacology and computational drug repositioning.

Polypharmacology is based on the well-known fact that many drugs act on several targets rather than on a single target [22]. A recent approach, which used network analysis to predict polypharmacology, involved the connection of 3700 FDA-approved or investigational drugs to their respective binding targets based on their activity profiles [23]. The aim of this work was to find novel off-target effects of existing drugs. Once the network was constructed, it could be interrogated using the 2D chemical structures of known drugs. For a given drug query, a score was calculated that represented the structural similarity between the query compound and the array of compounds displayed as binding a target. A statistical model was then applied to determine the significance of this overall structural similarity and targets exceeding a chosen significance threshold were predicted as having probable association with the query drug. This approach identified several novel polypharmacology relationships and one such prediction was confirmed using an *in vivo* knockout mouse model. It predicted several compounds that were not structurally related to the binding pocket of the target and would thus have been missed by rational drug design. In many cases, complex diseases may be the result of multiple pathways behaving aberrantly. In this scenario, one may wish to query such drug target networks by targets that have been independently linked to the same disease but that have no obvious connection at the pathway level. Missing data is a problem in this analysis as several compounds may not have been systematically tested across all assays of interest.

The main area where drug function networks can benefit drug discovery is polypharmacology and this could lead to drug repositioning [24]. Drug repositioning can lead to shorter development times particularly in cases where a drug has passed safety tests but may have failed due to lack of clinical efficacy. This can lead to significant reductions in drug discovery and development costs. Doing this computationally provides a way to systematically carry out an initial screening cheaply and quickly. In a recent study utilizing computational analysis of public gene expression data, Dudley *et al.* predicted that topiramate, an anticonvulsant drug, had the potential to treat a different type of disorder: inflammatory bowel disease (IBD) [25]. This was validated by demonstrating improvement in an IBD *in vivo* rodent model and highlights the potential of *in silico* approaches for discovering drug repositioning opportunities. In addition to drug target networks, several other

network analysis approaches have also been developed that are aimed at achieving this particular goal. Campillos *et al.* [26] constructed a side-effect network which modelled drugs and their side effects as nodes. The side-effect information for each drug was extracted from the respective drug package labels and translated into a standard vocabulary using a medical thesaurus. The authors used this approach to calculate drug–drug similarity coefficients based on their shared side-effect profiles and used these as edges to construct a network. These phenotypic side-effect similarities helped the authors infer cases where two drugs were likely to act through common targets. Compared to the drug target network approach described above, this network view provides a representation of resultant physiological consequences of drug activity. This approach has the potential to relate drugs to other drugs or diseases based on phenotype and can be beneficial when the molecular pathway resulting in the side effect is unknown. However, it may also be dangerous to infer what the exact molecular pathway is as many of the generally observed phenotypes (*e.g.* pain) may be the consequence of unrelated molecular processes. Other sources of phenotypic profiles that are available to model from, are the Mouse Genotype Database [27] and genome-wide association studies (GWAS; http://www.genome.gov/gwastudies) [28].

Another way to relate drugs to other drugs or diseases, for drug repositioning, is through the extent of their similarity or polarity in inducing cellular gene expression. Each drug or disease may be considered as inducing an array of specific gene expression changes in a cell. These changes may then be captured as a specific molecular expression profile representing the drug or disease. Disease-induced profiles can be considered as shifts from the equilibrium state of a cellular system and drugs that reverse this state towards equilibrium may represent therapeutic solutions. A public data resource that enables network analysis, based on this concept, is the Connectivity Map (cmap) project [29]. This project provides a reference collection of transcriptional response profiles for 1309 compounds that can be compared to disease expression profiles. Using the cmap data, Iorio *et al.* [30,31] constructed a drug–drug network using statistically significant transcriptional pairwise similarities between drugs as the edges. This helped them to identify 'communities' of drugs that act similarly at the molecular pathway level, often through their action against the same targets. Another application of the cmap data has been to construct drug–disease networks: Hu *et al.* [32] extracted disease-relevant expression data sets from the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) and used this to find drugs and diseases that clustered together by their gene expression profiles. This helped to suggest novel drug targets for diseases, whose relationships were not been previously known. However, a drawback of making inferences from cmap data on its own is that, until now, the data has been generated from a limited set of only five cell lines. This makes it difficult to extrapolate results to clinical phenotypes. One reason for this is that drugs often undergo physiologically induced changes before they reach their site of action. Furthermore, in most cases, only one concentration of the compound is tested against the cell line. Despite these limitations, the cmap approach can provide novel insights into new disease indications and compound mechanisms of action.

Novel links between drugs and human diseases can also be discovered by mining GWAS gene–disease associations for 'drugged' targets [28]. Mooser *et al.* hypothesised that if the disease indication of a drug target matches that of the GWAS disease, it increases confidence in the pursued indication. Conversely, unmatched (and hence previously unknown) disease indications represent drug repositioning opportunities. The group demonstrated the value of this approach by finding new disease indications for 92 (out of 155) gene targets.

Recently, Li *et al.* used a complex network analysis approach to identify relationships between diseases, genes and pathways [33]. The authors firstly used literature-mining to identify links between diseases and genes. This was followed by a statistical enrichment methodology to identify targets that share significant overlaps between diseases and pathways. This helped to propose pathways as therapeutic intervention points. As exemplified in this work, published literature is a rich source of interaction data. However, because of various groups using differing terminologies to refer to the same entity, automatically mined results still yield high false positive rates. Companies, based around natural language processing, for example Linguamatics (http://www.linguamatics.com/), provide ontologies (controlled vocabularies) of diseases, drugs and biological terms that could provide a starting point for this sort of analysis. The big challenge with ontologies is how to integrate and relate observations from one knowledge domain to another. To address this, there are concerted efforts to develop biomedical ontologies. Examples include the Open Biomedical Ontologies (OBO) consortium [34] and the development of the Chem2Bio2RDF semantic framework [35]. Problems arise when different groups use the same term to describe different concepts [36]. Data can also be sparse in certain domains, which will limit how much integration can be done, for example, around less-well studied disease areas.

Social networks

In contrast to biological and drug network analysis, investigation of social networks represents an overall different data realm, which can also aid drug companies. The pharmaceutical industry is facing major hurdles such as loss of exclusivity as branded products go off-patent [37] and increasing R&D budgets required to sustain a flat number of approved drugs coming on to the market each year [38]. It is, therefore, imperative for drug companies to find ways to increase the number of successful therapies transitioning through their pipelines.

One way the industry has traditionally sought to increase the number of new drugs being approved annually is through the establishment of mergers and acquisitions. However, an increasingly attractive alternative business model that drug companies are pursuing is to maximize their access to open innovation to deliver breakthrough therapies [39]. This can be done by fostering both external and internal ideas. Collaboration through public or private partnership can open up a spectrum of opportunities including pharma-academic consortia, joint ventures, spin-offs and in/out-licensing. Recent examples include the Pistoia Alliance (http://www.pistoiaalliance.org/) and the partnership between GSK and the McLaren Group (http://www.gsk.com/media/pressreleases/2011/2011-pressrelease-625498.htm).

Social networking is essential for pharmaceutical companies to forge new alliances with external institutes. The increased

visibility of global activities, made possible through the Internet, is removing communication barriers and facilitating the identification of both worldwide collaborations and business development opportunities. This is, therefore, an efficient resource to tap into to identify potential KOLs, leading institutes, innovative technologies in addition to providing a mechanism to monitor competitor activity. From a drug discovery perspective, social networking websites such as LinkedIn (http://linkedin.com/) aid in the establishment of relationships that could lead to new business opportunities. However, several other types of databases exist that can be queried computationally to identify social networking patterns. Examples include: (i) Orphanet (http://www.orpha.net/consor/cgi-bin/index.php) for highlighting leading expertise in rare diseases, (ii) SiteTrove (http://www.citeline.com/products/sitetrove) for establishing key clinicians working with specific patient populations and (iii) the peer-reviewed scientific literature for the identification of co-authors.

Querying these diverse resources generates lists of individuals or institutes that can become a bottleneck to prioritise. Network analysis provides a means to interpret these lists: a common approach may be to model the individuals as nodes and shared collaborations as edges. For example, Bulik-Sullivan and Sullivan [40] built a GWAS co-authorship network to visualise the collaborative nature behind GWAS studies dating between 2005 and 2010. The resulting scale-free network included 8718 individuals, represented as nodes, and 575,078 edges, which connected co-authors. Graph theory was applied onto the network so that questions such as which research groups were the most influential or which authors might be more inclined to drive collaborations could be elucidated. Several websites have adopted visualisation of social networks. Examples summarised in Fig. 1 include the LinkedIn InMaps, BioMedExperts and GoPubmed.

Following the initial network generation, citation counts and journal impact factors (for example as calculated by Thomson Reuters; http://thomsonreuters.com/products_services/science/science_products/a-z/science_citation_index/, http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/) serve as an optional criteria that can aid in further prioritisation. Once key authors have been identified, the subsequent step involves manually reviewing the underlying supporting evidence. This can be achieved by retrieving supporting articles from specialized science publication websites like Scirus (http://www.scirus.com/). This website enables retrieving journal content for authors and further highlights scientists' homepages, courseware, pre-print server material, patents and institutional website information.

There are of course challenges represented by this type of network analysis. For example, social networks can become complex and there is a danger with the notion that everyone is approximately six steps away from any other person. Networks can also become 'contaminated' if there is more than one individual with the same name. Tracking the author with associated affiliation details can help to disambiguate authors with matching names. Co-authors detailed in PubMed are associated with first author affiliations. Normalization Engine for Matching Organizations (NEMO), for example, has been designed to extract organisation names from PubMed affiliation strings and normalises them into canonical names which can be linked to the co-authors [41]. One issue with this is that not all co-authors may be from the same institute. Alternative efforts exist such as that offered by SciVerse's Scopus (http://www.info.sciverse.com/scopus/), a bibliographic database, to enable the systematic retrieval of specific contact details for each author cited in a user defined set of articles. An overview of the efforts in the field of author disambiguation is provided by Smalheiser and Torvik [42]. An alternative method to obtain current contact details is simply to carry out an Internet search of authors or institutes. Pursuing results from this technique may also be subject to social hurdles. For example, language barriers can arise if an institute's website is not easily translatable. Access to the level of contact details can also vary due to cultural practices: this may range from the availability of a detailed curriculum vitae to an email address (which may be out of date). A final challenge is the mechanism through which to contact potential collaborators. A fine balance needs to exist in the steps required to initiate dialogue with potential KOLs and to ensure that communications are not interpreted as spam.

Concluding remarks

In this article, we have highlighted a variety of network analysis approaches, which are being used to facilitate drug discovery. It has been suggested that both biological and social networks have 'scale-free' mathematical properties [43] indicating that their interactions are non-random. This highlights the potential to query networks to uncover previously unknown relationships. However, as we have discussed, there are still large gaps in our knowledge that prevent us from reconstructing the ultimate drug discovery network: the cell-to-physiology hierarchical relationship. To address this, biologists need to work closely and iteratively with computational scientists to test and refine network driven hypotheses.

Conflicts of interest

The authors declare that there is no conflict of interest associated with publishing the contents of this review article.

Acknowledgement

The authors would like to thank Philippe Sanseau, Computational Biology (GlaxoSmithKline), for editorial support in the development of this manuscript.

References

1 Terzer, M. *et al.* (2009) Genome-scale metabolic networks. *Wiley. Interdiscip. Rev. Syst. Biol. Med.* 1, 285–297

2 Beste, D.J. *et al.* (2007) GSMN-TB: a web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism. *Genome Biol.* 8, R89

3 Duarte, N.C. *et al.* (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1777–1782

4 Selvarasu, S. *et al.* (2012) Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnol. Bioeng.* 109, 1415–1429

5 Bonde, B.K. *et al.* (2011) Differential producibility analysis (DPA) of transcriptomic data with metabolic networks: deconstructing the metabolic response of *M. tuberculosis*. *PLoS Comput. Biol.* 7, E1002060
6 Li, Z. *et al.* (2011) Two-stage flux balance analysis of metabolic networks for drug target identification. *BMC Syst. Biol.* 5 (Suppl. 1), S11
7 Folger, O. *et al.* (2011) Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* 7, 501
8 Jamshidi, N. and Palsson, B.O. (2008) Formulating genome-scale kinetic models in the post-genome era. *Mol. Syst. Biol.* 4, 171
9 Pieroni, E. *et al.* (2008) Protein networking: insights into global functional organization of proteomes. *Proteomics* 8, 799–816
10 Aranda, B. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.* 38, D525–D531
11 Kelder, T. *et al.* (2012) WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.* 40, D1301–D1307
12 Croft, D. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 39, D691–D697
13 Navratil, V. *et al.* (2011) When the human viral infectome and diseasome networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC Syst. Biol.* 5, 13
14 Demir, E. *et al.* (2010) The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.* 28, 935–942
15 Nibbe, R.K. *et al.* (2010) An integrative -omics approach to identify functional sub-networks in human colorectal cancer. *PLoS Comput. Biol.* 6, E1000639
16 Chuang, H.Y. *et al.* (2007) Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* 3, 140
17 Saez-Rodriguez, J. *et al.* (2011) Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer Res.* 71, 5400–5411
18 Ideker, T. and Krogan, N.J. (2012) Differential network biology. *Mol. Syst. Biol.* 8, 565
19 Stumpf, M.P. *et al.* (2008) Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6959–6964
20 D'Eustachio, P. (2011) Reactome knowledgebase of human biological pathways and processes. *Methods Mol. Biol.* 694, 49–61
21 Chong, C.R. and Sullivan, D.J., Jr (2007) New uses for old drugs. *Nature* 448, 645–646
22 Achenbach, J. *et al.* (2011) Computational tools for polypharmacology and repurposing. *Future Med. Chem.* 3, 961–968
23 Keiser, M.J. *et al.* (2009) Predicting new molecular targets for known drugs. *Nature* 462, 175–181
24 Hopkins, A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* 4, 682–690
25 Dudley, J.T. *et al.* (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* 3, 96ra76
26 Campillos, M. *et al.* (2008) Drug target identification using side-effect similarity. *Science* 321, 263–266
27 Ringwald, M. and Eppig, J.T. (2011) Mouse mutants and phenotypes: accessing information for the study of mammalian gene function. *Methods* 53, 405–410
28 Sanseau, P. *et al.* (2012) Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.* 30, 317–320
29 Lamb, J. *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929–1935
30 Iorio, F. *et al.* (2010) Identification of small molecules enhancing autophagic function from drug network analysis. *Autophagy* 6, 1204–1205
31 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
32 Hu, G. and Agarwal, P. (2009) Human disease–drug network based on genomic expression profiles. *PLoS ONE* 4, E6536
33 Li, Y. and Agarwal, P. (2009) A pathway-based view of human diseases and disease relationships. *PLoS ONE* 4, E4346
34 Smith, B. *et al.* (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251–1255
35 Chen, B. *et al.* (2010) Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinform.* 11, 255
36 Rubin, D.L. *et al.* (2008) Biomedical ontologies: a functional perspective. *Brief Bioinform.* 9, 75–90
37 Goodman, M. (2009) Pharmaceutical industry financial performance. *Nat. Rev. Drug Discov.* 8, 927–928
38 Malik, N.N. (2009) Key issues in the pharmaceutical industry: consequences on R&D. *Expert Opin. Drug Discov.* 4, 15–19
39 Hunter, J. and Stephens, S. (2010) Is open innovation the way forward for big pharma? *Nat. Rev. Drug Discov.* 9, 87–88
40 Bulik-Sullivan, B.K. and Sullivan, P.F. (2012) The authorship network of genome-wide association studies. *Nat. Genet.* 44, 113
41 Jonnalagadda, S.R. and Topham, P. (2010) NEMO: extraction and normalization of organization names from PubMed affiliations. *J. Biomed. Discov. Collab.* 5, 50–75
42 Torvik, V.I. and Smalheiser, N.R. (2009) Author name disambiguation in MEDLINE. *ACM Trans. Knowl. Discov. Data* 3
43 Barabasi, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science* 286, 509–512
44 Kuhn, M. *et al.* (2010) STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res.* 38, D552–D556
45 Yildirim, M.A. *et al.* (2007) Drug-target network. *Nat. Biotechnol.* 25, 1119–1126
46 Hoehndorf, R. *et al.* (2011) Integrating systems biology models and biomedical ontologies. *BMC Syst. Biol.* 5, 124
47 Sirota, M. *et al.* (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci. Transl. Med.* 3, 96ra77
48 Goh, K.I. *et al.* (2007) The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8685–8690