Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Pharmacology

In silico methods to predict drug toxicity

Alessandra Roncaglioni, Andrey A Toropov, Alla P Toropova and Emilio Benfenati

This review describes *in silico* methods to characterize the toxicity of pharmaceuticals, including tools which predict toxicity endpoints such as genotoxicity or organ-specific models, tools addressing ADME processes, and methods focusing on protein–ligand docking binding.These *in silico* tools are rapidly evolving. Nowadays, the interest has shifted from classical studies to support toxicity screening of candidates, toward the use of *in silico* methods to support the expert. These methods, previously considered useful only to provide a rough, initial estimation, currently have attracted interest as they can assist the expert in investigating toxic potential. They provide the expert with safety perspectives and insights within a weight-of-evidence strategy. This represents a shift of the general philosophy of *in silico* methodology, and it is likely to further evolve especially exploiting links with system biology.

Addresses
IRCCS – Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milano, Italy

Corresponding author: Benfenati, Emilio (emilio.benfenati@marionegri.it, benfenati@marionegri.it)

Current Opinion in Pharmacology 2013, **13**:802–806

This review comes from a themed issue on **New technologies**

Edited by **Andreas Gescher** and **Marco Gobbi**

For a complete overview see the Issue and the Editorial

Available online 21st June 2013

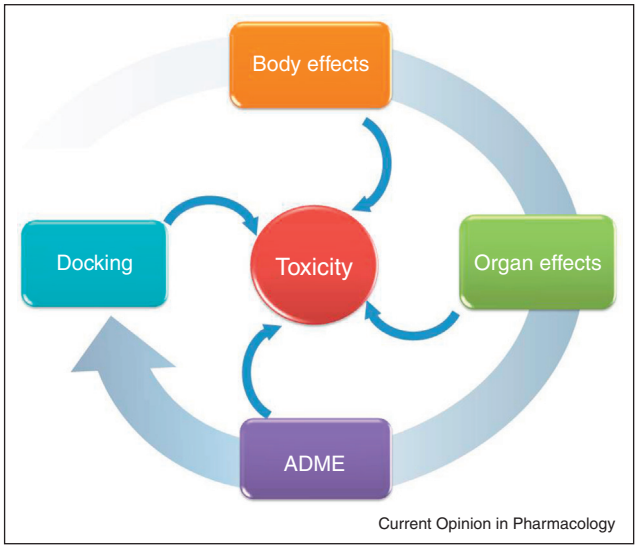
1471-4892/$ – see front matter,

http://dx.doi.org/10.1016/j.coph.2013.06.001

Introduction

The pharmaceutical industry has used *in silico* methods for decades to aid the search for new drugs. Usually the process of drug design implies the optimization of compounds to enhance activity through the identified therapeutic target starting with the screening of large sets of structures, using fast tools to identify an optimal number of promising lead compounds, and moving afterwards to more sophisticated tools which allow fine tuning the structures of the final candidates. It is nowadays well established that the rate of candidate drugs successfully placed into the market is very poor mainly due to the discovery of adverse effects. Therefore there is an increasing interest in early detection methods for finding possible reason for drug failure. *In silico* methods are particularly interesting from this point of view because they can be easily integrated into the early stage of the drug discovery process using only the 'virtual' structure of the compounds. Moreover, they are less time consuming and cheaper than wet experiments so that large numbers of compounds can be evaluated. There are several kinds of *in silico* models, which focus on targets at different levels, addressing the whole body, or specific organs, or certain biological processes, or focused biochemical mechanism, such as binding to a receptor (see Figure 1). A large set of *in silico* methodologies are used to predict toxicity. They include quantitative structure–activity relationship (QSAR) models, expert systems, 3D-QSAR and docking models. The QSAR methods seek a mathematical relationship between a group of molecular descriptors, used to describe each molecule present in a set of chemicals, and their toxicity values. It is fundamental to verify the predictivity of the model, using appropriate statistical methods. These methods include internal and external validation. In case of the external validation a set of new chemicals, never before used in model development, is applied. A recent book described the theory and applications of *in silico* models with emphasis on QSAR [1]. 'Expert systems' are software programs which codify a series of rules identified by experts in their field of interest. A typical example is when there is a set of known toxic fragments (often called 'structural alerts'), and the software recognizes their presence in the target chemical. An issue associated with this kind of approach is that the set of toxic fragments can be incomplete and thus may produce false negatives, that is, falsely predict chemicals as safe. 3D-QSAR is based on the concept of so-called 'molecular interaction fields'. With this technique the variation in the steric and electrostatic interaction energies calculated between each molecule and a probe is correlated with the variation in the investigated property. 3D-QSAR is usually focused on a set of structurally similar compounds with associated toxicity data, while docking mimics the binding to a biological macromolecule (usually a protein) [2]. In general, docking is not so commonly used in toxicity estimation, because most typically toxicity phenomena involve a complex sequence of events and binding to a specific receptor is only a possible component of this sequence. Conversely, docking studies are more frequently used in the drug design process where the therapeutic target is known (i.e. the receptor one wishes to block/activate), while often the causes are not clarified at the biochemical level for toxicological phenomena. This mini-review will present recent trends on the use of *in silico* methods to explore adverse effects of drugs. In

Figure 1



The more and more detailed tools available to predict toxicity endpoints, using *in silico* methods.

particular models and approaches dealing with ADME (adsorption, distribution, metabolism, and excretion) properties will be presented here, together with models related to toxicity, with emphasis on both systemic toxicity and organ specific effects. Furthermore we will discuss the applications of *in silico* methods from the regulatory point of view, and we consider problems and future perspectives.

Models for ADME

ADME properties comprise the evaluation of a series of parameters that are relevant not only in determining the capability of the drug to reach pharmacologically active concentration at the therapeutic targets. Furthermore, ADME may be responsible for adverse effects due for instance to accumulation or biotransformation. So ADME evaluation can be useful combined to *in silico* models for toxicity evaluation [3]. Physico-chemical indicators are increasingly used during the early stages of drug discovery to provide a comprehensive understanding of the key properties that affect biological functions (ADME). Historically simple approaches based on filters applied to the evaluation of some physico-chemical properties in relation to drug likeness (such as the 'Lipinski rule of 5') have been widely used and are still popular nowadays [4], although their use has been sometimes criticized for being over simplistic [5]. Lipinski rules are specific cut-offs values assigned to physico-chemical parameters (molecular weight, lipophilicity, number of H-bond donors or acceptors) to identify easily bioavailable drugs.

Drug bioavailability is a key aspect often evaluated with *in silico* models [6]. It comprises the assessment of common physico-chemical properties such as solubility [7] or pK_a, which describe the substance regardless of the biological environment. Moreover, interactions with the biological system are also evaluated, such as intestinal permeability [6] (with models derived on the basis of *in vivo* [8•] or *in vitro* data [9••]). Interest is now growing also in the role of active transporters [10] together with passive diffusion.

Some studies addressed the blood–brain barrier (BBB) permeability of compounds [11•,12] that can be related to possible neurotoxic effects. Binding to plasma proteins is also often evaluated [13].

Among the properties relevant to ADME, metabolism is one of the most intensely studied. In particular, metabolism can play a crucial role in the toxicity of drugs (e.g. genotoxicity; see also below: 'Regulatory context for *in silico* models') but also in drug–drug interactions. A variety of modeling approaches have been adopted in this field as recently detailed by Kirchmair *et al.* [14••]; these approaches include expert systems, data mining, QSARs, MIFs, and protein–ligand docking. Often, emphasis is placed on phase I metabolism involving P450 enzymes. Aspects which have been analyzed are selectivity for some CYP isoforms, metabolic product formation, relative quantities of metabolite formation, and sites of transformation in the molecule. In the case of metabolism, it is possible to rely on the structure of the enzymes involved in the biotransformation for the modeling. So the structure based approach is one of the most promising in analyzing metabolism [15] and has been successfully employed on a series of macromolecules relevant in the ADME processes such as phase II metabolizing sulfotransferases. Expert systems have been also widely employed with the advantage of allowing simultaneously evaluation of ADME and toxicity characteristics [16,17]. Software availability for the developed *in silico* models of ADME properties is also crucial for their routine use, and several commercial and non-commercial software programs for ADME properties estimation are available [18••] and new ones are constantly being developed [19]. A key characteristic of some of these software programs is the possibility to have trainable models where in house data are included in the training set of the prebuilt model to improve its performance for the specific compounds of interest.

An aspect that is often emphasized in evaluating model reliability is the lack of adequate quantity and quality of data to be modeled [20], so there is the need to make more experimental data available to modelers [18••,21]. Nowadays there are also more and more efforts to integrate ADME predictions into physiologically based pharmacokinetic/pharmacodynamic (PBPK/PD) models [22]. For instance, there are models predicting volume of distribution and other parameters which can be useful in the overall evaluation of the drug distribution in the

body. The combination of *in silico* and *in vitro* derived parameters can be useful to mimic *in vivo* behavior [23].

In silico models for toxicity

When dealing with *in silico* models for toxicity we can distinguish methods which address the overall toxic phenomenon, for instance carcinogenicity, and methods which address factors of the process leading to toxicity manifestations. Thus, some models address systemic toxicity while others focus on organ-specific toxicity. There are *in silico* models for a large number of toxicity phenomena [24,25•]. Among these genotoxicity and carcinogenicity, although very complex, are among the most extensively studied. However, the availability of a model is different from the reliability of the model prediction. Moreover the possibility to satisfactorily apply a model to a pharmaceutical of interest may depend on toxicological data availability for molecules chemically related to this entity. There are papers in which model performance is assessed for general chemicals [26,27], or for pharmaceuticals, with respect to carcinogenicity and mutagenicity [28••]. In contrast, there is a paucity of models for developmental or reproductive toxicity [18••]. In comparison with the models presented above, *in silico* models for organ-specific effects are generally focused on pharmaceuticals, since data availability is most abundant for drug-like compounds. Among them hepatotoxicity has been frequently investigated [29], and nowadays increasing interest is also placed on cardiotoxicity [30,31] and nephrotoxicity [32]. Limited models are available for neurotoxicity or other effects such as phospholipidosis [33]. Some investigations involved the use of adverse effect databases [34] or combinations of effects against several receptor targets [35], using integrated risk indices.

Regulatory context for *in silico* models

The use of *in silico* methods to estimate toxicity is solicited in different legislation in the EU such as those concerning chemicals or cosmetics. Their use in the pharmaceutical field is more related to the R&D of new drugs avoiding potential adverse effects. Furthermore, regulatory criteria have been defined for the use of *in silico* predictions in drug safety for the genotoxicity evaluation of drug impurities where the presence or absence of specific structural alerts triggers the subsequent management of potential risk posed by the presence of this impurity [36•,37,38•]. In this context, new models have been recently proposed to estimate genotoxic or carcinogenic potency in order to estimate safety [37,38•,39]. Very good results have been obtained in the evaluation of genotoxic impurities in the industrial sector when *in silico* data were coupled with expert evaluation [36•] achieving a negative predictive value of 99%.

Problems and future perspectives

As mentioned above, data availability is one of the major barriers to the improvement of *in silico* predictive models. Indeed, these methods are based on data, and frequently method viability suffers from a scarcity of data. In the case that interest focuses on drugs with a similar structures, the requirement of chemical entity number is low, typically 10–30. This number has to be much higher for heterogeneous compounds. A complication associated with unusual molecular scaffold features in new drug design is that such drugs may contain chemical moieties not present in the molecules used to build up the model generating effects difficult to predict. So far two types of solutions have been proposed to overcome this problem: firstly introduction of an independent third party entity allowing inclusion of confidential data into the database [40], and secondly the use of software capable of extracting rules that can be run by the owner of the data [41] and shared on a limited basis to avoid confidentiality issues. A similar problem is associated with data quality. The quality of the existing toxicological data and their standardization can be vital. A debate is ongoing on the definition of the different toxicity endpoints, on the similarity of the procedures and on data uncertainty [42].

New scientific fields will very probably have an impact on *in silico* methods for toxicity prediction. This is the case for systems biology, which provides associations useful for toxicity assessment, merging data sets of different origin [31,43] and contributing to the elucidation of mechanisms. For instance networks of protein-protein interactions have been investigated to elucidate adverse cardiotoxic effects [31]. This broad approach will add further complexity, on the one hand, but it will offer novel insights and possibilities to model toxicity phenomena, on the other.

Another challenge is that *in silico* models typically work on classical chemical structures, but nowadays new types of therapeutics emerge, including peptides and nanomaterials, therefore pioneer studies on peptides [44•] and nano-materials [45•,46] are a promising area of future expansion for *in silico* models. These studies addressed the issue of which format should be used to simplify the complexity of the parameters potentially involved, which range from the chemical structure to others experimentally measured.

Conclusions

In silico models for toxicity prediction are used within the pharmaceutical industry. The results obtained using these models can contribute to the screening process of new drugs. A wide series of different tools exist, addressing general toxicity effects, or organ-specific ones. Tools to predict ADME are common and assist in understanding toxicity phenomena.

The improvement in the field is expected when further experimental data will become available. Novel powerful approaches will help, integrating results obtained in

multiple independent ways. Indeed, as with *in vitro* methods, the combination of multiple approaches will increase the performance of the predictions, compared to results related to one model.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Benfenati E (Ed): **Theory, guidance and applications on QSAR and REACH**. E-book available at: http://www.orchestra-qsar.eu/documents/333.

2. Roncaglioni A, Benfenati E: **In silico-aided prediction of biological properties of chemicals: oestrogen receptor-mediated effects**. *Chem Soc Rev* 2008, **37**:441-450.

3. Modi S, Hughes M, Garrow A, White A: **The value of in silico chemistry in the safety assessment of chemicals in the consumer goods and pharmaceutical industries**. *Drug Discov Today* 2012, **17**:135-142.

4. Cox PB, Gregg RJ, Vasudevan A: **Abbott Physicochemical Tiering (APT)—a unified approach to HTS triage**. *Bioorg Med Chem* 2012, **20**:4564-4573.

5. Jacob A, Pratuangdejkul J, Buffet S, Launay J-M, Manivet P: **In silico platform for xenobiotics ADME-T pharmacological properties modeling and prediction. Part II: the body in a Hilbertian space**. *Drug Discov Today* 2009, **14**:406-412.

6. Hou T, Li Y, Zhang W, Wang J: **Recent developments of in silico predictions of intestinal absorption and oral bioavailability**. *Comb Chem High Throughput Screen* 2009, **12**:497-506.

7. Wang J, Hou T: **Recent advances on aqueous solubility prediction**. *Comb Chem High Throughput Screen* 2011, **14**:328-338.

8. • Talevi A, Goodarzi M, Ortiz EV, Duchowicz PR, Bellera CL, Pesce G, Castro EA, Bruno-Blanch LE: **Prediction of drug intestinal absorption by new linear and non-linear QSPR**. *Eur J Med Chem* 2011, **46**:218-228.
This papers shows examples of QSAR models for permeability. Non-linear models provide better results compared to linear ones.

9. •• Larregieu CA, Benet LZ: **Drug discovery and regulatory considerations for improving in silico and in vitro predictions that use Caco-2 as a surrogate for human intestinal permeability measurements**. *AAPS J* 2013, **15**:483-497.
It provides an interesting discussion on the example of the models for permeability, with possibilities and open issues; the last ones include the reproducibility of the experimental data, a typical problem for all in silico models, but also for any evaluation.

10. Bahadduri PM, Polli JE, Swaan PW, Ekins S: **Targeting drug transporters — combining in silico and in vitro approaches to predict in vivo**. *Methods Mol Biol* 2010, **637**:65-103.

11. • Martins IF, Teixeira AL, Pinheiro L, Falcao AO: **A Bayesian approach to in silico blood-brain barrier penetration modeling**. *J Chem Inf Model* 2012, **52**:1686-1697.
This article presents a new QSAR model for the blood brain barrier penetration, using a non-linear algorithm. The software is freely available on the internet.

12. Chen H, Winiwarter S, Fridén M, Antonsson M, Engkvist O: **In silico prediction of unbound brain-to-plasma concentration ratio using machine learning algorithms**. *J Mol Graph Model* 2011, **29**:985-995.

13. Li H, Chen Z, Xu X, Sui X, Guo T, Liu W, Zhang J: **Predicting human plasma protein binding of drugs using plasma protein interaction QSAR analysis (PPI-QSAR)**. *Biopharm Drug Dispos* 2011, **32**:333-342.

14. •• Kirchmair J, Williamson MJ, Tyzack JD, Tan L, Bond PJ, Bender A, Glen RC: **Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms**. *J Chem Inf Model* 2012, **52**:617-648.
A detailed overview of the *in silico* approaches for modeling metabolism, covering the different strategies, programs, and results.

15. Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA: **Toward in silico structure-based ADMET prediction in drug discovery**. *Drug Discov Today* 2012, **17**:44-55.

16. Valerio LG Jr, Long A: **The in silico prediction of human-specific metabolites from hepatotoxic drugs**. *Curr Drug Discov Technol* 2010, **7**:170-187.

17. Long A: **Drug metabolism in silico — the knowledge-based expert system approach. Historical perspectives and current strategies**. *Drug Discov Today* 2013, **10**:147-153.

18. •• Worth A, Fuart-Gatnik M, Lapenna S, Lo Piparo E, Mostrag-Szlichtyng A, Serafimova R: *The Use of Computational Methods in the Toxicological Assessment of Chemicals in Food: Current Status and Future Prospects, Scientific and Technical Research Series EUR 24748 EN.*. Luxembourg: Publications Office of the European Union; 2011, http://dx.doi.org/10.2788/6234.
An overview of the existing models to predict toxicity endpoints of different nature, including general toxicity endpoints and organ specific models.

19. Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, Lee PW, Tang Y: **admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties**. *J Chem Inf Model* 2012, **52**:3099-3105.

20. Kortagere S, Ekins S: **Troubleshooting computational methods in drug discovery**. *J Pharm Toxicol Methods* 2010, **61**:67-75.

21. Cao D, Wang J, Zhou R, Li Y, Yu H, Hou T: **ADMET evaluation in drug discovery, 11 PharmacoKinetics Knowledge Base (PKKB): a comprehensive database of pharmacokinetic and toxic properties for drugs**. *J Chem Inf Model* 2012, **52**:1132-1137.

22. Hosea NA, Jones HM: **Predicting pharmacokinetic profiles using in silico derived parameters**. *Mol Pharm* 2013, **10**:1207-1215.

23. Malmborg JJ, Ploeger BA: **Predicting human exposure of active drug after oral prodrug administration, using a joined in vitro/in silico–in vivo extrapolation and physiologically-based pharmacokinetic modeling approach**. *J Pharmacol Toxicol Methods* 2013, **67**:129-214.

24. Kar S, Kunal Roy K: **Predictive toxicology using QSAR: a perspective**. *J Indian Chem Soc* 2010, **87**:1455-1515.

25. • Kar S, Roy K: **Development and validation of a robust model for prediction of carcinogenicity of drugs**. *Indian J Biochem Biophys* 2011, **48**:111-122.
An example of QSAR model for carcinogenicity potency prediction, on the basis of a large data set. Most typically in silico models predict a carcinogenicity category.

26. Bakhtyari NG, Raitano G, Benfenati E, Martin T, Young D: **Comparison of in silico models for prediction of mutagenicity**. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* 2013, **31**:45-66.

27. Milan C, Schifanella O, Roncaglioni A, Benfenati E: **Comparison and possible use of in silico tools for carcinogenicity within REACH legislation**. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* 2011, **29**:300-323.

28. •• Hillebrecht A, Muster W, Brigo A, Kansy M, Weiser T, Singer T: **Comparative evaluation of in silico systems for ames test mutagenicity prediction: scope and limitations**. *Chem Res Toxicol* 2011, **24**:843-854.
This paper compares the results of different in silico methods for mutagenicity, considering drug-like compounds. In some cases good results can be obtained.

29. Diaz Ochoa JG, Bucher J, Péry ARR, Zaldivar Comenges JM, Niklas J, Mauch K: **A multi-scale modeling framework for individualized, spatiotemporal prediction of drug effects and toxicological risk**. *Front Pharmacol* 2013, **3**:204 http://dx.doi.org/10.3389/fphar.2012.00204.

30. Frid AA, Matthews EJ: **Prediction of drug-related cardiac adverse effects in humans-B: use of QSAR programs for early detection of drug-induced cardiac toxicities**. *Regul Toxicol Pharm* 2010, **56**:276-289.

31. Huang LC, Wu X, Chen JY: **Predicting adverse side effects of drugs**. *BMC Genomics* 2011, **12(Suppl. 5)**:S11 http://dx.doi.org/10.1186/1471-2164-12-S5-S11.

32. Myshkin E, Brennan R, Khasanova T, Sitnik T, Serebriyskaya T, Litvinova E, Guryanov A, Nikolsky Y, Nikolskaya T, Bureeva S: **Prediction of organ toxicity endpoints by QSAR modeling based on precise chemical-histopathology annotations**. *Chem Biol Drug Des* 2012, **80**:406-416.

33. Choi SS, Kim JS, Valerio LG Jr, Sadrieh N: **In silico modeling to predict drug-induced phospholipidosis**. *Toxicol Appl Pharmacol* 2013, **269**:195-204.

34. Cheng F, Li W, Wang X, Zhou Y, Wu Z, Shen J, Tang Y: **Adverse drug events: database construction and in silico prediction**. *J Chem Inf Model* 2013 http://dx.doi.org/10.1021/ci4000079. (March 22).

35. Vedani A, Dobler M, Smieško M: **VirtualToxLab — a platform for estimating the toxic potential of drugs, chemicals and natural products**. *Toxicol Appl Pharmacol* 2012, **261**:142-153.

36. • Dobo KL, Greene N, Fred C, Glowienke S, Harvey JS, Hasselgren C, Jolly R, Kenyon MO, Munzner JB, Muster W *et al.*: **In silico methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: An industry survey**. *Regul Toxicol Pharmacol* 2012, **62**:449-455.
This paper reports the results of a survey between different pharmaceutical industries, using in silico models for mutagenicity. Predictivity is good, and the use of expert judgment improves the reliability.

37. Contrera JF: **Improved in silico prediction of carcinogenic potency (TD50) and the risk specific dose (RSD) adjusted threshold of toxicological concern (TTC) for genotoxic chemicals and pharmaceutical impurities**. *Regul Toxicol Pharmacol* 2011, **59**:133-141.

38. • Valerio LG Jr, Cross KP: **Characterization and validation of an in silico toxicology model to predict the mutagenic potential of drug impurities**. *Toxicol Appl Pharmacol* 2012, **260**:209-221.
This papers shows results of the use of in silico models for mutagenicity, which can be useful in case of impurities of pharmaceuticals.

39. Bercu JP, Morton SM, Deahl JT, Gombar VK, Callis CM, van Lier RB: **In silico approaches to predicting cancer potency for risk assessment of genotoxic impurities in drug substances**. *Regul Toxicol Pharmacol* 2010, **57**:300-306.

40. Briggs K, Cases M, Heard DJ, Pastor M, Pognan F, Sanz F, Schwab CH, Steger-Hartmann T, Sutter A, Watson DK, Wichard JD: **Inroads to predict in vivo toxicology — an introduction to the eTOX project**. *Int J Mol Sci* 2012, **13**:3820-3846.

41. Ferrari T, Gini G, Bakhtyari NG, Benfenati E: **Mining toxicity structural alerts from SMILES: a new way to derive structure activity relationships**. In *Proceedings IEEE Symposium Series on Computational Intelligence*; *Paris, France, IEEE Computational Intelligence Society, Piscataway, NJ, April 11–15: 2011:120-127*.

42. Hardy B, Apic G, Carthew P, Clark D, Cook D, Dix I, Escher S, Hastings J, Heard DJ, Jeliazkova N *et al.*: **Toxicology ontology perspectives**. *ALTEX* 2012, **29**:139-156.

43. Zhou W, Huang C, Li Y, Duan J, Wang Y, Yang L: **A systematic identification of multiple toxin-target interactions based on chemical, genomic and toxicological data**. *Toxicology* 2012, **304**:173-184.

44. • Toropov AA, Toropova AP, Raska I Jr, Benfenati E, Gini G: **QSAR modeling of endpoints for peptides which is based on representation of the molecular structure by a sequence of amino acids**. *Struct Chem* 2012, **23**:1891-1904.
A novel, simplified QSAR approach for peptides is proposed, which accounts for the aminoacidic composition. The classical QSAR models do not address peptides.

45. • Toropov AA, Toropova AP, Benfenati E, Leszczynska D, Leszczynski J: **SMILES-based optimal descriptors: QSAR analysis of fullerene-based HIV-1 PR inhibitors by means of balance of correlations**. *J Comput Chem* 2010, **31**:381-392.
A free, new QSAR model for nanoparticles is described, with a program available through the internet, which can be applied to related cases.

46. Toropov AA, Toropova AP, Benfenati E, Gini G, Puzyn T, Leszczynska D, Leszczynski J: **Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria *Escherichia coli***. *Chemosphere* 2012, **89**:1098-1102.

The International Serious Adverse Events Consortium

Arthur L. Holden, Jorge L. Contreras, Sally John and Matthew R. Nelson

The International Serious Adverse Events Consortium is generating novel insights into the genetics and biology of drug-induced serious adverse events, and thereby improving pharmaceutical product development and decision-making.

Arthur L. Holden is Founder, CEO and Chairman of the iSAEC, 8770 West Bryn Mawr Avenue, Suite 1300, Chicago, Illinois 60631, USA.
Jorge L. Contreras is Legal Counsel to the iSAEC and at the S.J. Quinney College of Law, and Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA.
Sally John is Co-Chairperson of the iSAEC Scientific Management Committee and at Pfizer, New London, Connecticut 06320, USA.
Matthew R. Nelson is Co-Chairperson of the iSAEC Scientific Management Committee and at GlaxoSmithKline, Research Triangle Park, North Carolina 27709–3398, USA.
Correspondence to A.L.H.
e-mail: aholden@earthlink.net
doi:10.1038/nrd4441

The impetus for the International Serious Adverse Events Consortium (iSAEC) arose from a series of interviews in 2006 with senior research and development leaders of major pharmaceutical companies, exploring how to build on the success of the SNP Consortium[1] to identify additional, high-value genomic research areas in which to apply this highly effective cross-industry collaborative model. The interviewees assigned the highest priority to exploring the genetic basis of drug-induced, rare serious adverse events (SAEs). In May 2006, with staff at the US Food and Drug Administration (FDA), we conceptualized the structure for a private, international research consortium to explore the genetic contribution to drug-induced SAEs. It was felt the opportunities for applying genomic technologies to better understand this vital aspect of drug safety would benefit both drug development and regulatory oversight. Equally significant were the complexity, logistics, management, risks, and cost associated with such a research initiative. No single institution possessed the resources, sufficient well-phenotyped cases, genomics expertise and international breadth to execute such a research endeavour alone. The stage was set for the development and launch of the iSAEC.

Scientific focus and organizational structure

The iSAEC is a pharmaceutical-industry-led and FDA-supported international research consortium, focused on identifying and validating DNA variants predictive of the risk of drug-induced SAEs. It was launched in 2007 with the scientific and financial support of six funding members (Abbott, GlaxoSmithKline, Johnson & Johnson, Pfizer, Roche and Sanofi-Aventis). Additional dues-paying members were added (Novartis, Takeda, Daiichi Sankyo, and The Wellcome Trust) as the consortium completed its Phase 1 research programme (focused on the genetics of drug-induced liver injury (DILI) and serious skin injury (DISI)). A separate call for funding and membership roster was developed for the Phase 2 research programme, which included ten dues-paying members (Abbott, GlaxoSmithKline, Pfizer, Takeda, Daiichi Sankyo, Novartis, Merck, Amgen, AstraZeneca and the Wellcome Trust), as well as three associate members that made in-kind, non-cash contributions to the research effort (Cerner, Clinical Data and Catholic Health Initiatives). The FDA has participated from the outset as an observer, advisor and research collaborator, but without formal membership status.

Since 2007, the iSAEC has collaborated with over 200 leading academic centres and scientists globally to:

- standardize and publish phenotype definitions for the major drug-induced SAEs (liver, skin, heart and renal injury);
- build diverse, well-phenotyped clinical cohorts and sample repositories for many of the major SAEs;
- apply optimal genomic and computational methods (including imputation) for effective genome-wide single nucleotide polymorphism (SNP) genotyping and exome sequencing;
- ensure timely public availability of scientific results/associated data (within 12 months after genotyping, regardless of publication timing) to the scientific community at no cost[2] (see Further information); and
- ensure the open use of all iSAEC data, unencumbered by intellectual property constraints[3].

The iSAEC's organization is virtual and composed of multiple collaborative teams, staffed by member volunteers and research collaborators, and under the direction of the iSAEC's CEO/Chairman. The iSAEC is governed by a board of directors (BOD) that consists of one director from each sponsoring member and the CEO, *ex officio*, and makes its decisions using a 'majority rules' model. The board delegates the oversight and management of the consortium's research agenda to the scientific management committee (SMC), which has representatives from each member company as well as scientific and clinical experts from many of its major collaborations. The SMC is supported by the Data Analysis and Coordination Center (DACC) at Columbia University as well as a network of genotyping and sequencing partners. The DACC coordinates the aggregation, quality control, analysis and release of all research data; prior to public data release, no

consortium member or collaborator may use the data for any purpose other than the advancement of the consortium's research (that is, there is no preferential access; see Further information for details of the data release policy).

Current status and scientific output so far

Over the past 7 years, the iSAEC has developed novel, international clinical networks to aggregate well-phenotyped case collections associated with specific SAEs and causal drugs. Specifically, we have aggregated subjects with DILI, DISI, drug-induced hypersensitivity syndrome (DIHSS), drug-induced renal injury (DIRI), drug-induced Torsades de pointes/prolonged QT effects (DITdP), inflammatory bowel disease (IBD) therapy-related SAEs such as pancreatitis and leukopenia, excessive weight gain (EWG) associated with class 2 antipsychotics, and osteonecrosis of the jaw (ONJ). Case enrolment has been completed for all SAEs, with the exception of DIRI and those related to IBD (see Supplementary information S1 (table)). By the end of 2015, the consortium expects to have aggregated close to 7,500 SAE cases spanning these phenotypes. The majority of this collection will be Caucasian, but it will contain important African, Indian and Chinese cohorts. The scale, depth, quality, and diversity of this recruitment effort are unprecedented in the history of drug safety research.

The iSAEC has or will conduct genome-wide genotyping of all collected subjects. In Phase 1, initial genome-wide association studies were conducted for DILI, DISI and DITdP, leading to several novel findings and key insights into the primary immune-related mechanisms underlying many of these SAEs (see Supplementary information S2 (box) for a list of publications). Following the success of the first phase, the BOD approved a plan to increase the existing DILI and DISI case collections, expand into DIHSS, DIRI, EWG, ONJ and IBD-related SAEs, expand investigations for selected SAEs into non-European populations, and explore the role of rare variants in SAEs with pilot exome sequencing studies for co-amoxiclav-induced liver injury, clozapine-induced agranulocytosis and DITdP.

To date, the iSAEC has completed 18 public releases of anonymized subject-level clinical and genotyping data, associated with 3,623 of its cases and controls. A total of 135 researchers and institutions have applied for and been granted access to the iSAEC database (see Further information). Through this open access policy, we hope to stimulate further analysis that will yield additional scientific insights and publications as collections and genetic analysis methods evolve[2].

The iSAEC is helping to set the precedent for genetic analysis of drug-induced SAEs and beginning to broaden the scientific understanding of these highly personalized reactions to otherwise safe and effective drugs. Through our research, we have demonstrated that the primary genetic contribution to SAE risk is through human leukocyte antigen (HLA) variation and the adaptive immune response, and that the variants with clinically meaningful effects can be detected in relatively small sample sizes (<50 cases in several instances). This bodes well for the feasibility of applying genomic methods in the future when an immunologically mediated toxicity is suspected. In those studies where we have performed sequencing analysis, our quest to identify rare variants (that is, <1% of the population) with a large SAE influence has, to date, been unfruitful. We remain uncertain as to the effects such rare genetic variants may have on SAEs. To date, most of our findings are drug-specific versus across multiple drugs, which may be expected given the important role for the major histocompatibility complex genomic region in the pathology of immunologically mediated SAEs and the very specific relationships observed between HLA alleles and clinical disease (for example, HLA-B*27 in ankylosing spondylitis and HLA-C*06 in psoriasis). Finally, there are a number of HLA alleles that are associated with different SAEs and for different drugs, including HLA-B*57:01, HLA-DRB1*07:01, and HLA-DRB1*15:01, that may provide important insights into the underlying biology of SAEs and offer strategies to predict or mitigate future SAEs.

Lessons learned and conclusions

Lessons learned in developing the iSAEC include:

- a clear, unifying, highly important mission is a must from the outset;
- to maximize membership and ease of formation, ensure the proposed effort is precompetitive and in the public good;
- develop the operating plan and uniform membership requirements with the potential funding members;
- have a high-quality, phased scientific/operating plan before recruiting funding members;
- establish dedicated, high-quality management early;
- develop funding requirements early, and work with the potential members on trade-offs to produce an affordable and effectively phased consortium;
- organize a board and well-defined committees with high-quality, dedicated leaders;
- outsource to the best external advisors/investigators via performance-based contracts;
- exceed expectations;
- make it fun and say "thank you" in meaningful ways;
- know when to terminate the consortium — begin with the end in mind!

Drug-related biomedical research options are exploding in number, complexity, risk and cost. To address the challenges, all stakeholders must work together to develop new collaborative research frameworks and diversified funding models that enhance financial leverage and research productivity. The iSAEC serves as an excellent example of such innovation.

1. Holden, A. L. The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques* **32**, S22–S24 (2002).
2. Contreras, J. L., Floratos, A. & Holden, A.L. The International Serious Adverse Events Consortium's data sharing model. *Nature Biotech.* **31**, 17–19 (2012).
3. Contreras, J. L. Bermuda's legacy: policy, patents, and the design of the genome commons. *Minn. J. Law Sci. Technol.* **12**, 61–125 (2011).

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
iSAEC Data Access Site: https://dataportal.saeconsortium.org/
iSAEC Public Data Access Policy: http://www.saeconsortium.org/?q=node/27

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (box)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Science Translational Medicine AAAS

Predicting Adverse Drug Events Using Pharmacological Network Models
Aurel Cami, *et al.*
Sci Transl Med **3**, 114ra127 (2011);
DOI: 10.1126/scitranslmed.3002774

Editor's Summary

The Power of Prediction

We've all done it: googled a combination of medical terms to describe how we feel after taking a new medication. The result is a seemingly infinite list of Web sites telling us that the nausea is normal, or that the headaches warrant another visit to the doctor. Oftentimes, important adverse effects of drugs are discovered and added to the drug label only years after a drug goes on the market. But what if scientists could know about certain adverse drug effects before they are clinically discovered? Cami and colleagues develop a mathematical approach to predicting such adverse events associated with the drugs we take, in hopes of reducing drug-related morbidity —and mortality.

After its release to the market, any given drug undergoes rigorous evaluation to determine associated ADEs (adverse drug effects). This post hoc analysis is usually unable to detect rare or delayed-onset ADEs until enough clinical evidence accumulates–a process that may take years. The method devised by Cami and coauthors does not need to wait for such evidence to accumulate. Instead, it can inform drug safety practitioners early on of likely ADEs that will be detected down the line.

The authors first collected a "snapshot" of 809 drugs and their 852 related adverse events that had been documented in 2005. These drug-safety associations were combined with taxonomic and biological data to construct a network that is reminiscent of a web. Cami *et al.* then used this drug-ADE network to train a logistic regression predictive model —basically creating a formula that would indicate the likelihood of unknown side effects of any drug in the network. The predictive capabilities of the model were prospectively validated using drug-ADE associations newly reported between 2006 and 2010. Such prospective evaluation preserves the chronological order of drug adverse event reporting, making it a realistic method for predicting future ADEs. With their network, the authors were able to predict with high specificity seven of eight drug ADEs identified by pharmacological experts as having emerged after 2005, including the relationship between the anti-diabetes drug rosiglitazone (Avandia) and heart attack.

The benefit for patients? With this powerful model in place, certain unknown adverse drug effects may be discovered earlier, helping to prevent drug-related morbidity and mortality through appropriate consumer label warnings.

Science Translational Medicine (print ISSN 1946-6234; online ISSN 1946-6242) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

COMPUTATIONAL PHARMACOLOGY

Predicting Adverse Drug Events Using Pharmacological Network Models

Aurel Cami,[1,2]* Alana Arnold,[1] Shannon Manzi,[1] Ben Reis[1,2]

Early and accurate identification of adverse drug events (ADEs) is critically important for public health. We have developed a novel approach for predicting ADEs, called predictive pharmacosafety networks (PPNs). PPNs integrate the network structure formed by known drug-ADE relationships with information on specific drugs and adverse events to predict likely unknown ADEs. Rather than waiting for sufficient post-market evidence to accumulate for a given ADE, this predictive approach relies on leveraging existing, contextual drug safety information, thereby having the potential to identify certain ADEs earlier. We constructed a network representation of drug-ADE associations for 809 drugs and 852 ADEs on the basis of a snapshot of a widely used drug safety database from 2005 and supplemented these data with additional pharmacological information. We trained a logistic regression model to predict unknown drug-ADE associations that were not listed in the 2005 snapshot. We evaluated the model's performance by comparing these predictions with the new drug-ADE associations that appeared in a 2010 snapshot of the same drug safety database. The proposed model achieved an AUROC (area under the receiver operating characteristic curve) statistic of 0.87, with a sensitivity of 0.42 given a specificity of 0.95. These findings suggest that predictive network methods can be useful for predicting unknown ADEs.

INTRODUCTION

Adverse drug events (ADEs) pose serious challenges to public health. A wide range of approaches are used in attempts to detect ADEs in both pre- and post-market stages. In pre-market stages, compounds undergo extensive toxicity testing through a diverse range of methods (*1*–*11*), followed by rigorous clinical trials to evaluate their efficacy and safety profile. In post-market stages, data collected in different types of observational databases—such as spontaneous reports, drug-specific patient registries, administrative claims databases, and electronic health records—are continuously analyzed in search of evidence of increased ADE rates possibly related to specific drugs (*12*–*14*). These methods involve well-known limitations, such as the difficulty of detecting rare or delayed-onset ADEs (*15*–*17*), as well as ADEs that are already common in the treatment population (*18*).

To further strengthen the collection of tools available to drug safety professionals, we proposed an approach for identifying ADEs, called predictive pharmacosafety networks (PPN). PPNs exploit the network structure formed by known drug safety relationships and combine this high-level network data with information on specific drugs and adverse events to predict likely unknown adverse events. Although existing methods (*12*–*14*) rely on sufficient post-market evidence to accumulate for a specific adverse drug effect, this predictive approach relies on leveraging contextual information from previously known drug safety relationships. As a result, it has the potential to predict certain candidate ADEs earlier than they might be detected by existing post-market methods. In addition, the proposed approach can potentially become a valuable complementary addition to the set of existing pre-market predictive tools of toxicity through enabling the analysis of an unexplored type of information: the network structure.

We are aware of only one other network-based study focused on the prediction of unknown ADEs (*19*). In relation to this previous work, the current paper makes two primary contributions. First, it develops a predictive approach that integrates various data types—including structural network properties, drug intrinsic properties, and drug and ADE taxonomies—and introduces several covariates that have not been explored previously. Second, and perhaps more importantly, this study evaluated network-based predictive models through a simulated prospective approach. Prospective evaluation is the only method that preserves the chronological order in which the information historically became available. Preserving this chronological order is crucial because, to be useful in practice, an ADE prediction method must be able to predict an unknown ADE based only on information available before the ADE became known. The use of other approaches, such as cross-validation, can potentially break this chronological order by including in the training set ADEs that historically became known only after certain related ADEs in the validation set became known, providing an unfair and unrealistic advantage to the prediction model.

Earlier identification of ADEs could have a direct positive impact on drug safety and public health by enabling the design of risk evaluation and minimization strategies and the addition of appropriate label warnings and market withdrawals. The proposed computational approach is intended as a complementary hypothesis generation tool to identify potential drug–adverse event relationships. Given the tremendous complexity of drug safety research and the many ways in which a drug can cause an adverse event (*20*, *21*), each of these potential relationships requires further investigation, including expert human pharmacological review.

RESULTS

Overview of constructing PPNs

As the first stage of the PPN approach (Fig. 1), we integrated data from multiple sources, including data on drug-ADE associations, drug

[1]Children's Hospital Boston, Boston, MA 02115, USA. [2]Harvard Medical School, Harvard University, Boston, MA 02115, USA.
*To whom correspondence should be addressed. E-mail: aurel.cami@childrens.harvard.edu

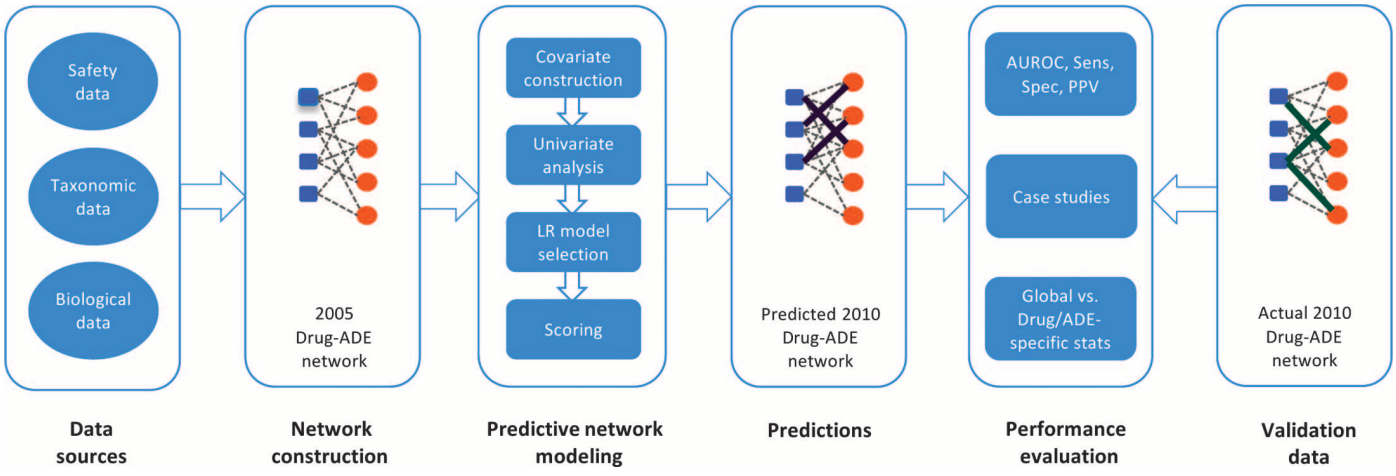


Fig. 1. Overview of the PPN approach. First, data were integrated from multiple sources, including safety data (drug-ADE associations contained in two chronologically separated snapshots of a drug safety database: one from 2005 and another from 2010), taxonomic data (ATC taxonomy of drugs and MedDRA taxonomy of ADEs), and biological data (intrinsic drug properties). Next, a network representation of the drug-ADE associations contained in the 2005 database snapshot was constructed. The drug-ADE network was used to derive a collection of network, taxonomic, and intrinsic covariates and to train a logistic regression (LR) predictive model. The trained LR model was used to generate scores for all the non-edges of the 2005 network. Next, the actual drug-ADE associations that were newly reported during the 5-year period from 2006 to 2010 were identified by comparing the 2005 snapshot with the 2010 snapshot. Finally, a systematic evaluation of the model's performance was performed by comparing the model-generated scores with the set of newly reported drug-ADE associations.

and ADE taxonomies, and intrinsic drug properties. Next, we constructed a network representation of the drug-ADE associations contained in the 2005 database snapshot. We used this drug-ADE network to derive a collection of network, taxonomic, and intrinsic covariates (table S1) and to train a logistic regression (LR) predictive model. The LR model generated scores (predicted probabilities) for all the non-edges of the 2005 network, and the highest-scoring non-edges formed the model's predictions for unknown drug-ADE associations. Next, we identified the drug-ADE associations that were newly reported during the 5-year period from 2006 to 2010 by comparing the 2005 and 2010 snapshots. We performed a systematic evaluation of the model's performance by comparing the model-generated scores with the set of newly reported drug-ADE associations.

Data description

Figure 2 provides a visualization of the drug-ADE network (high-resolution image available as supplementary online file "Fig2-highres.tif"). The data for constructing this network were obtained from different sources. Drug-ADE associations were extracted from two chronologically separated snapshots of a proprietary commercial database widely used in hospitals today, provided by Lexicomp (http://www.lexi.com). These two snapshots contained all reported adverse events of all Food and Drug Administration–approved drugs as of December 2005 and December 2010, respectively. The ADEs in these snapshots were then mapped to the Medical Dictionary for Regulatory Activities (MedDRA) taxonomy v12.0 (Supplementary Methods). For a very small number of ADE names (less than 1%), we were not able to find a mapping at the MedDRA lowest-level term (LLT), preferred term (PT), or high-level term (HLT) levels, but only at a higher level, such as high-level group terms (HLGTs). We excluded those ADE names from our analysis. The taxonomic and intrinsic drug properties were extracted from the following publicly available databases: World Health Organization Anatomical Therapeutic Chemical (ATC) Classification System (http://www.whocc.no/atc), University of Alberta DrugBank (http://www.drugbank.ca), and National Center for Biotechnology Information's PubChem Compound (http://www.ncbi.nlm.nih.gov/pccompound). Generic names were used to uniquely represent drugs and to perform data integration.

After integrating the safety data from the 2005 snapshot with the DrugBank, PubChem, and ATC drug data, we identified 809 unique drugs common to all three databases (table S2). These 809 drugs collectively were associated with 852 unique ADEs, represented by MedDRA HLTs, in the 2005 snapshot. The 2005 drug-ADE network therefore consisted of 1661 nodes (Fig. 2): 809 drugs and 852 ADEs. This network had 39,591 edges and 649,677 non-edges (proportion of edges in the training set: 5.7%). In the 2010 safety data snapshot, we identified 10,845 new edges between these 809 drugs and 852 ADEs (proportion of new edges in the validation set: 1.7%). Of the 809 drugs included in the study, 522 had at least one "newly associated" ADE that appeared as associated with the drug in the 2010 data, but not in the 2005 data. Figure S1 shows the mean number of newly associated ADEs per drug in each ATC top-level group. Over this subset of 522 drugs, the minimum, maximum, and mean number of newly associated ADEs were 1, 164, and 20.8 [95% confidence interval (CI), 18.7 to 22.9], respectively. Of the 852 ADEs included in the study, 709 had at least one newly associated drug that appeared as associated with the ADE in the 2010 data, but not in the 2005 data. Figure S2 shows the mean number of newly associated drugs per ADE in each MedDRA top-level group. Over this subset of 709 ADEs, the minimum, maximum, and mean number of newly associated drugs were 1, 97, and 15.3 (95% CI, 13.97 to 16.63), respectively.

Baseline model: Network covariates

We began our modeling analysis by first considering baseline models that contained only network covariates (those relating to the structure of connections between drugs and ADEs). We investigated nine

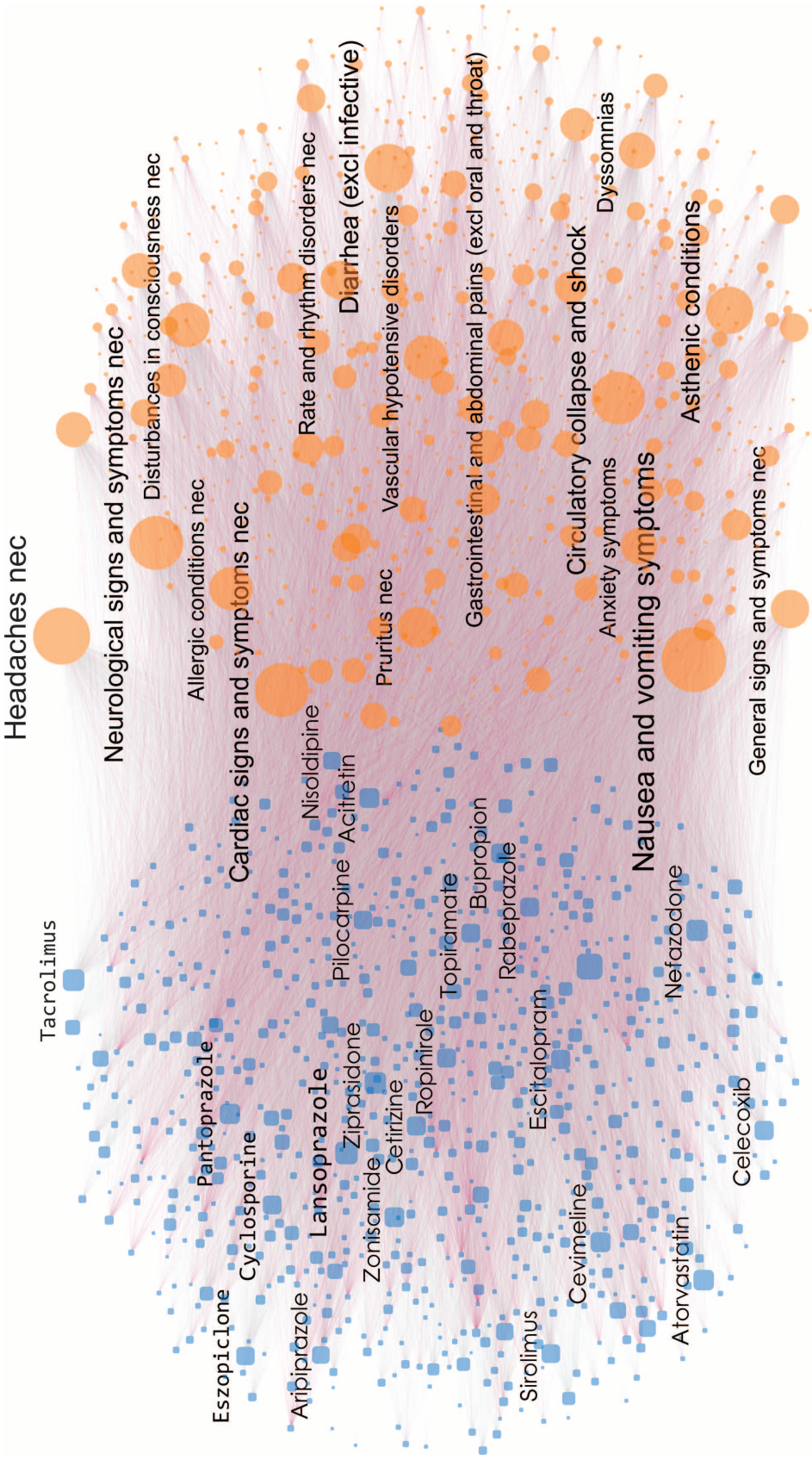


Fig. 2. A visualization of the drug-ADE network produced with the software package Cytoscape (http://www.cytoscape.org). The drug and ADE partitions are shown separated. The size of each node and node label is proportional to the node degree in the 2005 network. The edges of the 2005 network are shown in gray, whereas the edges newly introduced between 2006 and 2010 are shown in purple. A few drugs and ADEs have been labeled for illustrative purposes; the labeling of all nodes was not possible because of the commercial nature of the drug-ADE data used in this study. An expandable high-resolution version of this figure is available as the supplementary online file "Fig2-highres.tif."

Table 1. Analysis of model covariates. Definition of each covariate is provided in table S1. Parameter estimates were obtained by fitting univariate logistic regression models. AUROC values were determined on both the training data and the validation data and are therefore shown separately. Sen99, Sen95, and Sen90 denote the sensitivity achieved when specificity is fixed at 0.99, 0.95, and 0.9, respectively. Model NET contains all network covariates except degree-sum and degree-ratio. Model TAX contains all taxonomic covariates. Model INT contains all intrinsic covariates.

Covariate name	Parameter estimate	Training set AUROC	Validation set AUROC	Sen99	Sen95	Sen90
jackard-ADE-max	9.99	0.95	0.856	0.108	0.368	0.554
degree-prod	0.00023	0.918	0.853	0.118	0.374	0.555
degree-sum	0.013	0.906	0.828	0.11	0.363	0.507
jackard-drug-max	13.911	0.906	0.786	0.064	0.232	0.386
jackard-ADE-KL	−2.36	0.898	0.832	0.086	0.309	0.491
jackard-drug-KL	−2.07	0.854	0.771	0.066	0.249	0.401
degree-ratio	−0.181	0.799	0.781	0.032	0.196	0.363
degree-absdiff	0.01	0.772	0.688	0.093	0.267	0.386
NET		0.963	0.862	0.111	0.407	0.588
atc-min	−0.568	0.87	0.816	0	0	0.422
atc-KL	−2.85	0.759	0.72	0.06	0.215	0.365
meddra-min	−0.695	0.737	0.644	0	0	0
meddra-KL	−4.77	0.64	0.571	0.02	0.11	0.174
TAX		0.90	0.838	0.093	0.337	0.519
NET + TAX		0.965	0.869	0.119	0.415	0.608
euclid-min	−0.0044	0.815	0.752	0.048	0.2	0.365
euclid-KL	−1.6	0.839	0.79	0.077	0.294	0.458
INT		0.839	0.79	0.077	0.294	0.458
NET+INT		0.963	0.86	0.111	0.408	0.586
TAX + INT		0.907	0.844	0.107	0.346	0.525
NET + TAX + INT		0.966	0.869	0.122	0.421	0.612

network covariates (table S1): degree-prod, degree-sum, degree-ratio, degree-absdiff, jackard-ADE-max, jackard-ADE-Kullback-Leibler (KL) divergence, jackard-drug-max, jackard-drug-KL, and edge-density. For a given drug-ADE pair (i, j), each covariate $X_s(i, j)$ in this study depended on the nodes (i, j) and on their neighbor sets $N(i)$ and $N(j)$. The fitting of univariate and multivariate models was performed using the LOGISTIC procedure in the Statistical Analysis System (SAS) v9.2. The covariate degree-prod aimed to capture any potential preferential association among high-degree drugs and ADEs. The covariate degree-absdiff aimed to capture assortativity by degree; that is, whether high-degree drugs tend to connect more frequently to high-degree ADEs or to small-degree ADEs. The covariates degree-sum and degree-ratio are analogous to degree-prod and degree-absdiff, respectively, and were generated for completeness. Degree-based covariates were found to be predictive of drug-ADE associations and drug-drug interactions using an exponential random graph (ERG) model. The covariates jackard-ADE-max and jackard-drug-max aimed to capture the structural similarity between drug pairs and ADE pairs. Jackard coefficient–based predictors have been used in several earlier studies (*22*, *23*); specifically, jackard-drug-max was earlier used in (*19*). The versions of Jackard-based predictors based on KL divergence leverage the full distribution of similarities between a drug and the drugs in its local neighborhood or between an ADE and the ADEs in its neighborhood.

Univariate and multivariate analyses were performed for these network covariates (Table 1). The signs of the parameter estimates reflect the effect of each network covariate on the probability of the existence of an edge between a given drug and ADE. Specifically, edges are more likely to exist in drug-ADE pairs with higher values of degree-prod, degree-absdiff, jackard-drug-max, and jackard-ADE-max, and lower values of jackard-drug-KL and jackard-ADE-KL. Figure 3 illustrates the effect of covariates degree-prod (Fig. 3A) and jackard-ADE-max (Fig. 3B) on the probability of the existence of an edge for both the training and the validation sets. As expected from the differing proportions of edges in the training and validation sets (5.7% versus 1.7%, respectively), the probabilities were markedly smaller in the validation set. The mean distribution of the variable jackard-ADE $(J(k, j), k \in N(i) - \{j\})$ over the edges and non-edges groups, for both the training (Fig. 3F) and the validation (Fig. 3G) sets, is also shown, along with analogous information for the variable jackard-drug $(J(k, i), k \in N(j) - \{i\})$ (Fig. 3, H and I). It may be seen that when a drug-ADE pair (i, j) denotes a true association, the neighborhood $N(i)$ typically contains more ADEs that are highly similar—such as with a Jackard coefficient greater than 0.2—to the ADE j than when the pair denotes a non-association. An analogous conclusion can be drawn for neighborhoods $N(j)$.

The performance of the best model developed using only these network (NET) covariates was evaluated (Table 1). Table 2 lists the 10 highest-scoring true positives predicted with model NET.

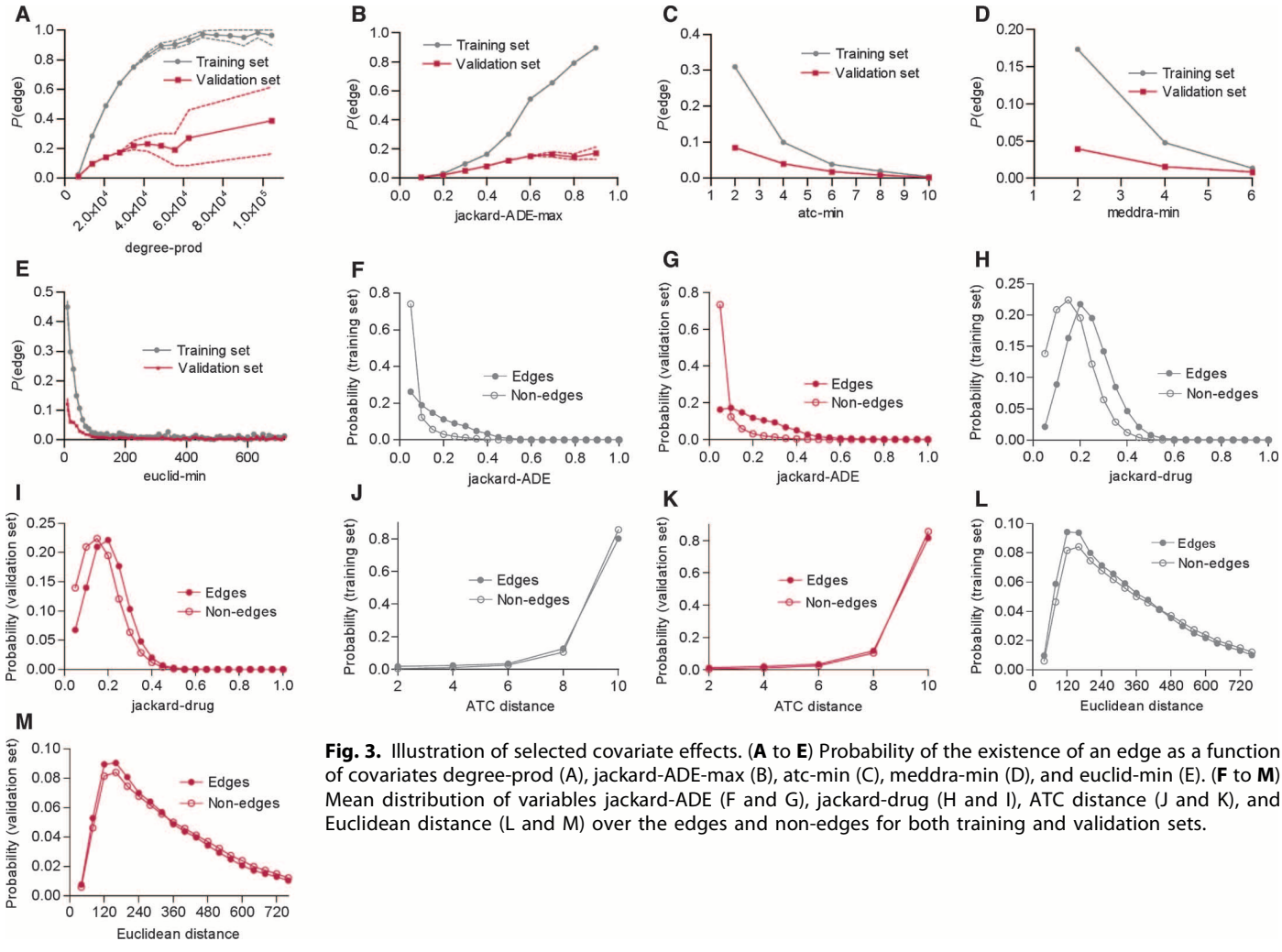


Fig. 3. Illustration of selected covariate effects. (**A** to **E**) Probability of the existence of an edge as a function of covariates degree-prod (A), jackard-ADE-max (B), atc-min (C), meddra-min (D), and euclid-min (E). (**F** to **M**) Mean distribution of variables jackard-ADE (F and G), jackard-drug (H and I), ATC distance (J and K), and Euclidean distance (L and M) over the edges and non-edges for both training and validation sets.

For comparison, if the prediction threshold were chosen such that the pairs in Table 2 were the only true positives, the model would predict 30 false positives, resulting in a positive predictive value (PPV) of 0.25.

Adding taxonomic covariates in the model

Next, we investigated the use of taxonomic covariates in the model. These covariates were based on the ATC taxonomy of drugs and the MedDRA taxonomy of adverse events. As a preliminary step, we computed for every pair ($drug_1$, $drug_2$) the minimum distance $d_{ATC}(drug_1, drug_2)$, denoting the minimum over all possible ATC positions of $drug_1$ and $drug_2$ of the length of the shortest path between $drug_1$ and $drug_2$ in the ATC taxonomy. Similarly, we computed for every pair of adverse events (ADE_1, ADE_2) the distance $d_{MedDRA}(ADE_1, ADE_2)$, denoting the length of the shortest path between ADE_1 and ADE_2 in the MedDRA taxonomy. Using the above distance measures, we constructed four taxonomic covariates: atc-min, atc-KL, meddra-min, and meddra-KL (table S1). ATC distance–based covariates were found to be predictive of drug-ADE associations and drug-drug interactions using an ERG model. For reference, Perlman *et al.* used a different ATC-based metric to predict drug targets (*24*). The MedDRA-based covariates defined above are counterparts of the ATC-based covariates. The motivation behind atc-KL and meddra-KL was the same as the motivation behind jackard-ADE-KL, discussed earlier.

Table 1 shows the results of univariate and multivariate analysis for the taxonomic covariates. It can be seen that edges are more likely to exist in drug-ADE pairs with smaller values of atc-min, meddra-min, atc-KL, and meddra-KL. Figure 3 illustrates the effect of atc-min (Fig. 3C) and meddra-min (Fig. 3D) on the probability of edge. The means of the distribution of ATC distances over the edges and non-edges groups are also shown for both the training (Fig. 3J) and the validation (Fig. 3K) sets. It may be seen that when a drug-ADE pair (i, j) denotes a true association, the neighborhood $N(j)$ typically contains more drugs that are at a small minimum ATC distance—such as at a distance of 2—to the drug i than when the pair denotes a non-association. The results in Table 1 show that a model containing all taxonomic (TAX) covariates performs reasonably well [validation set AUROC (area under the receiver operating characteristic curve), 0.838], but less well than the network (NET) model (validation set AUROC, 0.862). A model combining network and taxonomic covariates (NET + TAX) has a better performance (validation set AUROC, 0.869) than the network-only or the taxonomy-only models.

Table 2. The 10 highest-scoring true positives predicted with the NET model and their corresponding network covariates. Network covariates are specified in table S1. nec, not elsewhere classified.

Drug name	ADE	degree-prod	jackard-drug-max	jackard-ADE-max	jackard-drug-KL	jackard-ADE-KL	degree-absdiff
Nortriptyline	Circulatory collapse and shock	40,052	0.80	0.93	0.10	0.39	451
Fenoprofen	Edema nec	19,899	0.77	0.90	0.20	0.48	230
Ketoprofen	Edema nec	22,275	0.74	0.90	0.22	0.57	222
Nortriptyline	Headaches nec	43,624	0.80	0.74	0.10	0.37	498
Oxybutynin	Pupillary signs	2,862	0.50	0.95	0.40	1.02	1
Nabumetone	Edema nec	24,948	0.54	0.90	0.13	0.43	213
Hydrocortisone	Muscle weakness conditions	3,410	0.63	0.81	0.91	1.04	7
Prochlorperazine	Sexual arousal disorders	8,052	0.58	0.84	0.35	0.62	56
Metoprolol	Circulatory collapse and shock	22,661	0.59	0.93	0.23	0.90	484
Ibuprofen	Genitourinary tract infections and inflammations nec	5,124	0.69	0.72	0.61	0.78	23

Table 3. The 10 highest-scoring true positives predicted with the TAX model and their corresponding taxonomic covariates. Taxonomic covariates are specified in table S1.

Drug name	ADE	atc-min	meddra-min	atc-KL	meddra-KL
Triamcinolone	Hyperglycemic conditions nec	2	2	0.079	0.020
Cefdinir	Red blood cell analyses	2	2	0.088	0.007
Clomipramine	Purpura and related conditions	2	2	0.086	0.006
Indomethacin	Urinary abnormalities	2	2	0.093	0.007
Anagrelide	Skin injuries and mechanical dermatoses	2	2	0.038	0.012
Estramustine	Skin injuries and mechanical dermatoses	2	2	0.038	0.016
Betaxolol	Dyspneas	2	2	0.068	0.015
Isocarboxazid	Sleep disorders nec	2	2	0.137	0.025
Rivastigmine	Cerebrovascular and spinal necrosis and vascular insufficiency	2	2	0.063	0.011
Indomethacin	Hyperglycemic conditions nec	2	2	0.069	0.003

Table 3 lists the 10 highest-scoring true positives predicted with the TAX model. For comparison, if the prediction threshold were chosen such that the pairs in Table 3 were the only true positives, the model would predict 22 false positives, resulting in a PPV of 0.31.

Adding intrinsic covariates in the model

Finally, we investigated the use of intrinsic covariates in the model. For this purpose, we first assembled a vector of intrinsic properties for each drug. We extracted 16 drug molecular descriptors from PubChem: Molecular Weight, XLogP3 (partition coefficient), H Bond Donor, H Bond Acceptor, Rotatable Bond Count, Tautomer Count, Topological Polar Surface Area, Heavy Atom Count, Formal Charge, Complexity, Defined Atom StereoCenter (SC) Count, Undefined Atom SC Count, Defined Bond SC Count, Undefined Bond SC Count, Covalently Bonded (CB) Unit Count, and Isotope Atom Count. We excluded from our analysis XLogP3 and Tautomer Count because of missing values (table S3). We also excluded Isotope Atom Count because it had a value of zero for every drug in the study. Next, we extracted four physicochemical or absorption, distribution, metabolism, and excretion (ADME) properties from DrugBank: Melting Point, Exp LogP Hydrophobicity, Protein Binding, and Half Life. All properties extracted from DrugBank had a significant proportion of missing data (table S4). Here, we decided to analyze only the two DrugBank properties having the smallest proportions of missing data: Exp LogP Hydrophobicity and Protein Binding. For these two properties, we replaced the missing values by the average of non-missing values.

As a preliminary step, we computed for every pair ($drug_1$, $drug_2$) the Euclidean distance $d_{INT}(drug_1, drug_2)$ in the 15-dimensional intrinsic property space described above. Using this distance, we constructed two intrinsic covariates: euclid-min and euclid-KL (table S1).

Table 4. The 10 highest-scoring true positives predicted with the INT model and their corresponding intrinsic covariates. Intrinsic covariates are specified in table S1.

Drug name	ADE	euclid-min	euclid-KL
Fenoprofen	Febrile disorders	13.50358895	0.023528
Dexmedetomidine	Diarrhea*	12.02562272	0.023577
Benzphetamine	Diarrhea*	30.53129051	0.026821
Orphenadrine	Pruritus nec	20.73166319	0.028765
Dexmedetomidine	Gastrointestinal and abdominal pains (excl oral and throat)	12.02562272	0.029129
Diethylpropion	Diarrhea*	23.07050289	0.032107
Pramipexole	Diarrhea*	22.27254264	0.033017
Ropivacaine	Febrile disorders	30.79868175	0.033470
Dexmedetomidine	Headaches nec	12.02562272	0.033204
Nabumetone	Edema nec	25.31841108	0.033915

*Diarrhea ADEs exclude infective.

Covariates based on distance in the intrinsic attribute space were found to be predictive of drug-drug interactions using an ERG model. A distance measure similar to d_{INT} has been used to guide the clustering process in (*25*). The motivation behind the KL distance version of Euclidean distance is the same as the motivation discussed earlier for network and taxonomic covariates.

The results of univariate analyses for the intrinsic covariates show that edges are more likely to exist in drug-ADE pairs with smaller values of euclid-min and euclid-KL (Table 1). Figure 3 further illustrates the effect of euclid-min (Fig. 3E) on the probability of the existence of an edge. The means of the distributions of Euclidean distances in the neighborhood of drug-ADE pairs over the edges and non-edges groups, for both the training (Fig. 3L) and the validation (Fig. 3M) sets, are also shown. It may be seen that when a drug-ADE pair (i, j) denotes a true association, the neighborhood $N(j)$ typically contains more drugs that are at a small Euclidean distance—such as a distance less than 200—to the drug i than when the pair denotes a non-association.

A model containing only intrinsic (INT) covariates performs less well than the network or taxonomic models (validation set AUROC, 0.79) (Table 1). Adding intrinsic covariates to network or taxonomic models leads to a slight or no improvement in performance. For example, TAX + INT model reaches a slightly greater AUROC of 0.844 compared to the AUROC of 0.838 reached by TAX model, whereas NET + TAX + INT model reaches the same AUROC as does NET + TAX, but slightly higher sensitivity at the specificity levels 0.90, 0.95, and 0.99. To investigate the sensitivity of the predictive performance to the two physical/ADME properties that had missing and thus imputed data values (Exp LogP Hydrophobicity and Protein Binding), we repeated the performance evaluation with these two descriptors excluded. We found that the validation set AUROCs of covariates euclid-min and euclid-KL remained virtually unchanged (that is, 0.75 and 0.79, respectively) relative to the pre-exclusion AUROC values.

Table 4 lists the 10 highest-scoring true positives predicted with the model INT. If the prediction threshold were chosen such that the pairs in Table 4 were the only true positives, the model would predict 38 false positives, resulting in a PPV of 0.21. Intercorrelation of covariates was generally small (table S5): The highest positive (Pearson) correlation was between degree-prod and jackard-ADE-max (0.71), whereas the highest negative correlation was between jackard-ADE-KL and jackard-ADE-max (−0.77).

Comparison of model types

Comparative histograms of the scores generated by the three model types, for the observed edges and non-edges groups, were created (fig. S3). For all three models, the non-edges consistently obtained low scores, resulting in small (<0.2) thresholds for the 0.9, 0.95, and 0.99 specificity levels.

Three-way Venn diagrams for the sets of true positives and false positives were generated by the three model types, with specificity fixed at 0.95 (fig. S4). There are substantial differences in the correct predictions made by the three model types. However, each model also generated a number of distinct false positives. This may explain why the NET + TAX + INT model does not perform much better than the best single-type model.

Finally, comparative histograms are provided to illustrate the characteristics of the drug-ADE pairs that were predicted to be edges and non-edges by the models NET (fig. S5), TAX (fig. S6), and INT (fig. S7), when specificity of each model was fixed at 95%.

Drug- and ADE-specific prediction accuracy

The predictive performance of the models was evaluated on the set of all non-edges in the 2005 network through a comparison with the newly reported (in 2010) true drug-ADE associations. This straightforward approach allowed for a consistent and systematic evaluation of the model performance across 649,677 drug-ADE pairs. To investigate any potential variation in predictive performance according to drug and ADE, we carried out a second evaluation. For this evaluation, we generated all 809 drug-specific validation sets and 852 ADE-specific validation sets and, for each set, computed an AUROC statistic on the basis of the scores generated with the NET + TAX + INT model. However, because these drug- and ADE-specific AUROCs were based on local thresholds, they are not directly comparable to the AUROCs reported in Table 1, which were based on global thresholds.

Drug-specific AUROCs were plotted, with drugs grouped according to the ATC top-level categories (fig. S8A). For most of the drugs in the network, the AUROC was above 0.85 and mean AUROCs did not vary much across drug ATC categories. A plot of the drug-specific AUROC against the number of newly associated ADEs is also provided (fig. S8B). The P value associated with the slope of the regression line for this relationship was 0.86, implying that the AUROC was not affected significantly by the number of newly associated ADEs. However, the variability of AUROC was much higher when the number of newly associated ADEs was small than when this number was high. There are numerous drugs for which the NET + TAX + INT model produced a very high AUROC, including desipramine (0.977), chlorothiazide (0.958), and niacin (0.933). There are also a small number of drugs for which the NET + TAX + INT model produced a small AUROC, including cysteamine (0.599), cevimeline (0.601), and pioglitazone (0.637).

ADE-specific AUROCs were also plotted for MedDRA top-level groups and newly associated drugs (fig. S9). Examples of ADEs for which the NET + TAX + INT model produced high AUROCs include myelodysplastic syndromes (0.932), esophageal ulcers and perforation (0.928), and mood alterations with manic symptoms (0.909). Examples

of ADEs for which the NET + TAX + INT model did not produce high AUROCs include dermal and epidermal conditions not elsewhere classified (0.526), pulmonary edemas (0.562), and bladder infections and inflammations (0.549).

Prediction case studies

Two pharmacological experts were asked to choose a set of prominent drug-ADE associations that were discovered during the period from 2006 to 2010 for use as case studies. For each case study they named, we generated a score using the model NET + TAX + INT and then computed the specificity and PPV corresponding to that score (using the drug-specific validation sets). The respective values of score, specificity, and PPV obtained for the selected case studies are provided in table S6. The specificities corresponding to the model-generated scores were consistently high: All drug-ADE pairs shown, except "Saquinavir-Electrocardiogram QT prolonged," would have been detected using operational thresholds that correspond to a specificity level of 0.9. The PPV typically lies between 0.2 and 0.4, but in extreme cases, it is as low as 0.04 and as high as 0.67 (table S6), depending on the number of newly reported ADEs for the drug under consideration. That is, the greater the number of newly reported ADEs, the greater the PPV, as expected.

DISCUSSION

Here, we investigated the utility of network methods for predicting unknown drug adverse events in a simulated prospective setting. We trained a network-based predictive model on safety data from 2005 and used it to predict the new associations among 809 drugs and 852 ADEs reported between 2006 and 2010. The proposed model achieved an AUROC of 0.87, with a sensitivity of 0.42 (0.61) at a specificity of 0.95 (0.9). Our findings suggest that predictive network methods can be useful for predicting future reported drug-ADE relationships. The proposed approach, based on computer models and available data, can complement existing hypothesis generation tools that support the work of drug safety professionals.

The integrative data representation and the predictive method discussed here fall within the scope of the emerging field of systems pharmacology (*26*). In recent years, systems pharmacologic approaches have been applied successfully to various problems, such as identifying new targets for existing drugs (*23*, *24*, *27*–*30*) or understanding ADEs (*21*, *25*, *31*–*35*). Here, we extend our earlier work on developing network-based methods for predicting drug adverse events and drug-drug interactions. We had previously constructed integrated network representations of drug-drug interactions and ADEs and developed ERG (*36*) predictive models to identify the most likely "missing" edges (*22*, *37*) in those networks. Here, we explored the following directions. First, we developed more complex network-based models that incorporate additional types of data, such as simple molecular descriptors of drugs, and additional predictors, including novel ones like the KL-based covariates (table S1). To integrate these additional predictors in our model, we used a maximum-likelihood fitting method based on LR rather than the Markov chain Monte Carlo fitting method (*38*) used previously. We also evaluated the predictive models developed here through a simulated prospective study design that used two chronologically separated snapshots of a drug safety database.

We are aware of only one other study that attempts to predict unknown drug ADEs through network-based methods (*19*). That study and the current study are similar in that they both integrate various types of information to predict unknown likely ADEs and both conclude that prediction of reported ADEs may be possible. The data and methods used by the two studies differ in several ways. The current paper uses three types of information—network, intrinsic, and taxonomic—whereas Atias and Sharan (*19*) used a subset of the network information in the current study (jackard-drug-max), used alternative intrinsic information (hashed fingerprints computed from molecule structures), and did not include taxonomic information. Moreover, the predictive model described here takes into account all possible drug-ADE relationships at a given time (for example, 2005), yet (*19*) excludes from analysis any drugs and ADEs that have a degree smaller than two or that lie in the top 10% in terms of degree. Finally, perhaps the most important distinction between these two studies lies in the approach for evaluating the prediction accuracy. We used a simulated prospective approach, compared with (*19*), which used cross-validation on ADE data from the Side Effects Resource (SIDER, http://sideeffects.embl.de/) database and comparison of predictions with unseen ADE associations found in another database, the Hazardous Substances Data Bank (http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB)—neither of which preserves the chronological order in which the information became available.

The goal of our study was to demonstrate the utility of the proposed network-based approach, not to develop an optimal model or to discover an optimal set of predictors for a given model, which are much broader objectives. The set of covariates analyzed here was restricted to those network covariates that were relatively inexpensive computationally and those drug and ADE properties that were accessible. The covariates based on the drug intrinsic properties that we examined displayed a lower predictive performance than the network and taxonomic covariates. One potential explanation for these results is that here we have used only a limited set of basic intrinsic properties. However, a large amount of additional biological data exists—such as quantitative structure-activity relationship (QSAR) molecular descriptors (*39*)—that could be integrated into the PPN modeling framework to potentially improve the predictive performance. Varied performance by drug and ADE was found for the final predictive model. This variability may depend on many factors, such as the varying amount of information available for each drug and ADE, the existence of potential biases in the reporting of ADEs, and the limited ability of the model to correctly predict certain types of drug-ADE associations.

The gold standard used in this study was the set of all reported drug-ADE associations from the Lexicomp database mentioned above. Although this gold standard is the best one available, several limitations warrant discussion. First, we only had knowledge of the drug-ADE pairs that were reported to be true associations but not of the entirety of drug-ADE pairs that were tested and found unrelated. Thus, we do not know if the reported negatives were indeed true negatives or not yet known positives. In addition, a fraction of the reported drug-ADE associations may be false positives that were reported for various reasons, such as to protect against liability in cases of uncertainty. The strong performance of some study covariates, such as degree-prod, could reflect the fact that more testing and monitoring is conducted for certain drug-ADE pairs than others. Alternatively, it could be due to an inherent link between those covariates and the likelihood of true

drug-ADE associations. In the former case, the predictive value of these covariates may change in time if the underlying testing and monitoring processes also change. In the latter case, the predictive value of these covariates would remain unchanged. In practical terms, the reported ADEs are the ones used by the medical community today for clinical practice, and the early prediction of any of them could reduce drug-related morbidity and mortality. In the gold standard used in this paper, the reported drug-ADE associations constituted only 1.7% of all possible drug-ADE pairs. As is often the case with predicting rare phenomena, high specificity and sensitivity can be associated with a low PPV. However, this predictive approach can help drug safety professionals greatly reduce the vast space of possible drug-ADE pairs and better prioritize the use of their already limited resources.

This study could be extended in several clinically relevant directions. First, a number of statistical approaches may be used to deal with potential correlations in the response data, including ERG models (*36*, *38*) or mixed models. Furthermore, the network data could be stratified by various criteria, such as by ADE type or by drug-ADE mechanism, leading to a more accurate prediction of certain types of ADEs. Finally, the network data can be enriched with frequency information on drug-ADE associations. All these extensions could increase the practical value of the PPN approach for drug safety professionals.

In summary, we have proposed a new network-based method for predicting ADEs. This method can be applied to data that are available today and can be used as a practical hypothesis generation tool with potential to reduce the morbidity and mortality resulting from drug-related adverse events. A practitioner wishing to apply this method in a prospective "real-world" setting can access the latest version of any of the clinical drug safety databases available and use it to train the model described above. The model will generate a set of predicted drug-ADE relationships that do not appear in the training database. The practitioner can then follow up on the highest-scoring predictions with thorough clinical investigation, thus focusing his or her already limited time and resources on higher-probability drug safety hypotheses.

MATERIALS AND METHODS

Network construction

We constructed a bipartite network to represent the data on drugs, ADEs, and their associations. In this network, nodes denote drugs or ADEs and edges denote known drug-ADE associations. The set of edges corresponds to the drug-ADE associations contained in a 2005 snapshot of a drug safety database provided by Lexicomp. Because a single ADE name can map to multiple HLT codes, we replaced each drug-ADE association appearing in the original data snapshots with one or more drug-HLT associations, as determined by the MedDRA taxonomy. In addition, for each drug in the network, we assembled a vector that contains the drug's intrinsic properties extracted from DrugBank and PubChem as well as the drug's ATC code(s). We refer to the network described above as the drug-ADE network.

The predictive model

We model the binary response variable Y_{ij}, $i = 1, \ldots,$ number of drugs, $j = 1, \ldots,$ number of ADEs, denoting the presence or absence of drug-ADE associations. Using LR, we modeled this response as a Bernoulli random variable with expectation $E[Y_{ij}] = p_{ij}$, where p_{ij} is given by Eq. 1:

$$p_{ij} = 1/[1 + \exp(-\textstyle\sum_s q_s X_s(i, j))] \quad (1)$$

Here, q_s denotes the model parameter and X_s denotes the model covariate. The covariates used in this study are of three main types (table S1). Network covariates of the first type depend on the structure of the observed drug-ADE network but not on the attributes of drugs or ADEs. Taxonomic covariates of the second type depend on the structure of the observed drug-ADE network and on the taxonomic attributes (ATC and MedDRA codes). Intrinsic covariates of the third type depend on the structure of the observed drug-ADE network and on the intrinsic properties of drugs. The source code to compute all covariates is provided as supplementary online files: meddra_mapping_codes.sas, NET_INT_covariates.R, and TAX_covariates.sas (listed in table S7). We proceeded by making the standard assumption of independence between the responses Y_{ij} and carrying out model fitting by maximum likelihood.

Model prediction

The training data in our experiments consisted of the 2005 drug-ADE network. The validation data consisted of all the pairs in 2010 data that were non-edges in 2005. The prediction goal was to identify which of the pairs in the validation set would appear as edges in the 2010 snapshot of the drug safety database.

In the training phase, we used the binary response Y_{ij} corresponding to the 2005 drug-ADE network and analyzed a collection of network, taxonomic, and intrinsic covariates. For each collection of covariates, we performed univariate and multivariate analyses. In the univariate analysis stage, we fitted all possible univariate models, ranked the covariates on the basis of the training set AUROC, and excluded the bottom-ranked covariates from further analysis. In the multivariate analysis stage, we performed covariate selection through an exhaustive search over all possible sets of predictors using Akaike information criterion (AIC) goodness-of-fit measure to perform model ranking. The statistical significance (P values) of covariates was assessed through the standard χ^2 test in the LOGISTIC procedure in SAS.

After a multivariate model was trained, we scored each drug-ADE pair (i, j) in the validation set using the predicted probabilities pest_{ij}, as shown in Eq. 2:

$$\text{pest}_{ij} = 1/[1 + \exp(-\textstyle\sum_s \text{qest}_s X_s(i, j))] \quad (2)$$

We hypothesized that the validation set drug-ADE pairs having the highest predicted probabilities would be the ones that appear as true associations in the 2010 snapshot. To evaluate the predictive performance, we computed the validation set AUROC by comparing the scores given in Eq. 2 with the actual drug-ADE associations appearing in the 2010 snapshot. In addition, we computed the model sensitivity and PPV for various benchmark levels of specificity, including 0.99, 0.95, and 0.90.

SUPPLEMENTARY MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/3/114/114ra127/DC1

Methods

Table S1. Definition of covariates.

Table S2. List of drugs and their ATC codes.

Table S3. Number of missing observations for PubChem properties extracted for this study.

Table S4. Number of missing observations for DrugBank properties extracted for this study.
Table S5. Intercorrelation analysis of covariates.
Table S6. Prediction cases studies.
Table S7. List of supplementary source code files.
Fig. S1. Newly associated ADEs per drug in each ATC top-level group.
Fig. S2. Newly associated drugs per ADE in each MedDRA top-level group.
Fig. S3. Comparative histograms of scores for the observed edges and non-edges by the three model types.
Fig. S4. Three-way Venn diagrams for the sets of true and false positives generated by models NET, TAX, and INT.
Fig. S5. Comparative histograms of selected network covariates for the predicted edges and non-edges.
Fig. S6. Comparative histograms of selected taxonomic covariates for the predicted edges and non-edges.
Fig. S7. Comparative histograms of the intrinsic covariates for the predicted edges and non-edges.
Fig. S8. Drug-specific AUROCs.
Fig. S9. ADE-specific AUROCs.
File "meddra_mapping_code.sas" (SAS code to perform MedDRA mapping).
File "NET_INT_covariates.R" (R code to compute network and intrinsic covariates).
File "TAX_covariates.sas"(SAS code to compute taxonomic covariates).
File "Fig2-highres.tif" (high-resolution version of Fig. 2).

REFERENCES AND NOTES

1. J. C. Dearden, In silico prediction of drug toxicity. *J. Comput. Aided Mol. Des.* **17**, 119–127 (2003).
2. S. Ekins, J. Mestres, B. Testa, In silico pharmacology for drug discovery: Methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* **152**, 9–20 (2007).
3. J. Hamon, S. Whitebread, V. Techer-Etienne, H. Le Coq, K. Azzaoui, L. Urban, In vitro safety pharmacology profiling: What else beyond hERG? *Future Med. Chem.* **1**, 645–665 (2009).
4. M. R. Fieldent, K. L. Kolaja, The role of early in vivo toxicity testing in drug discovery toxicology. *Expert Opin. Drug Saf.* **7**, 107–110 (2008).
5. J. A. Kramer, J. E. Sagartz, D. L. Morris, The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. *Nat. Rev. Drug Discov.* **6**, 636–649 (2007).
6. C. J. Gross, J. A. Kramer, The role of investigative molecular toxicology in early stage drug development. *Expert Opin. Drug Saf.* **2**, 147–159 (2003).
7. A. P. Chiang, A. J. Butte, Data-driven methods to discover molecular determinants of serious adverse drug events. *Clin. Pharmacol. Ther.* **85**, 259–268 (2009).
8. J. L. Walgren, D. C. Thompson, Application of proteomic technologies in the drug development process. *Toxicol. Lett.* **149**, 377–385 (2004).
9. J. K. Nicholson, J. Connelly, J. C. Lindon, E. Holmes, Metabonomics: A platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **1**, 153–161 (2002).
10. W. H. Heijne, A. S. Kienhuis, B. van Ommen, R. H. Stierum, J. P. Groten, Systems toxicology: Applications of toxicogenomics, transcriptomics, proteomics and metabolomics in toxicology. *Expert Rev. Proteomics* **2**, 767–780 (2005).
11. E. C. Butcher, Can cell systems biology rescue drug discovery? *Nat. Rev. Drug Discov.* **4**, 461–467 (2005).
12. J. S. Almenoff, K. K. LaCroix, N. A. Yuen, D. Fram, W. DuMouchel, Comparative performance of two quantitative safety signalling methods: Implications for use in a pharmacovigilance department. *Drug Saf.* **29**, 875–887 (2006).
13. J. S. Brown, M. Kulldorff, K. A. Chan, R. L. Davis, D. Graham, P. T. Pettus, S. E. Andrade, M. A. Raebel, L. Herrinton, D. Roblin, D. Boudreau, D. Smith, J. H. Gurwitz, M. J. Gunter, R. Platt, Early detection of adverse drug events within population-based health networks: Application of sequential testing methods. *Pharmacoepidemiol. Drug Saf.* **16**, 1275–1284 (2007).
14. G. N. Norén, I. R. Edwards, Modern methods of pharmacovigilance: Detecting adverse effects of drugs. *Clin. Med.* **9**, 486–489 (2009).
15. C. Ellis, FDA alerts asthmatics to drug safety risk. *Nat. Rev. Drug Discov.* **2**, 765 (2003).
16. S. Frantz, Pharma's year of trouble and strife. *Nat. Rev. Drug Discov.* **5**, 7–9 (2006).
17. B. Hughes, 2007: Spotlight on drug safety. *Nat. Rev. Drug Discov.* **7**, 5–7 (2008).
18. A. Breckenridge, K. Woods, J. Raine, Monitoring the safety of licensed medicines. *Nat. Rev. Drug Discov.* **4**, 541–543 (2005).
19. N. Atias, R. Sharan, An algorithmic framework for predicting side-effects of drugs. *Res. Comp. Mol. Biol.* **6044**, 1–14 (2010).
20. R. E. Ferner, J. K. Aronson, EIDOS: A mechanistic classification of adverse drug effects. *Drug Saf.* **33**, 15–23 (2010).
21. S. I. Berger, R. Iyengar, Role of systems pharmacology in understanding drug adverse events. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **3**, 129–135 (2011).
22. D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks. *J. Am. Soc. Inform. Sci. Tech.* **58**, 1019–1031 (2007).
23. M. Campillos, M. Kuhn, A. C. Gavin, L. J. Jensen, P. Bork, Drug target identification using side-effect similarity. *Science* **321**, 263–266 (2008).
24. L. Perlman, A. Gottlieb, N. Atias, E. Ruppin, R. Sharan, Combining drug and gene similarity measures for drug-target elucidation. *J. Comput. Biol.* **18**, 133–145 (2011).
25. A. F. Fliri, W. T. Loging, P. F. Thadeio, R. A. Volkmann, Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat. Chem. Biol.* **1**, 389–397 (2005).
26. A. L. Hopkins, Network pharmacology: The next paradigm in drug discovery. *Nat. Chem. Biol.* **4**, 682–690 (2008).
27. M. Kuhn, M. Campillos, P. González, L. J. Jensen, P. Bork, Large-scale prediction of drug-target relationships. *FEBS Lett.* **582**, 1283–1290 (2008).
28. M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijer, R. C. Matos, T. B. Tran, R. Whaley, R. A. Glennon, J. Hert, K. L. Thomas, D. D. Edwards, B. K. Shoichet, B. L. Roth, Predicting new molecular targets for known drugs. *Nature* **462**, 175–181 (2009).
29. Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* **24**, i232–i240 (2008).
30. S. I. Berger, R. Iyengar, Network analyses in systems pharmacology. *Bioinformatics* **25**, 2466–2472 (2009).
31. A. F. Fliri, W. T. Loging, P. F. Thadeio, R. A. Volkmann, Biospectra analysis: Model proteome characterizations for linking molecular structure and biological response. *J. Med. Chem.* **48**, 6918–6925 (2005).
32. A. F. Fliri, W. T. Loging, R. A. Volkmann, Analysis of system structure–function relationships. *ChemMedChem* **2**, 1774–1782 (2007).
33. J. Scheiber, J. L. Jenkins, S. C. K. Sukuru, A. Bender, D. Mikhailov, M. Milik, K. Azzaoui, S. Whitebread, J. Hamon, L. Urban, M. Glick, J. W. Davies, Mapping adverse drug reactions in chemical space. *J. Med. Chem.* **52**, 3103–3107 (2009).
34. S. F. Lin, K. T. Xiao, Y. T. Huang, C. C. Chiu, V. W. Soo, Analysis of adverse drug reactions using drug and drug target interactions and graph-based methods. *Artif. Intell. Med.* **48**, 161–166 (2010).
35. G. Jin, X. Zhou, H. Wang, H. Zhao, K. Cui, X. S. Zhang, L. Chen, S. L. Hazen, K. Li, S. T. Wong, The knowledge-integrated network biomarkers discovery for major adverse cardiac events. *J. Proteome Res.* **7**, 4013–4021 (2008).
36. S. M. Goodreau, M. S. Handcock, D. R. Hunter, C. T. Butts, M. Morris, A statnet tutorial. *J. Stat. Softw.* **24**, 1–27 (2008).
37. A. Clauset, C. Moore, M. E. Newman, Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98–101 (2008).
38. D. R. Hunter, S. M. Goodreau, M. S. Handcock, Goodness of fit of social network models. *J. Am. Stat. Assoc.* **103**, 248–258 (2008).
39. D. Butina, M. D. Segall, K. Frankcombe, Predicting ADME properties in silico: Methods and models. *Drug Discov. Today* **7**, S83–S88 (2002).

Acknowledgments: We thank Lexicomp for providing the drug safety data and K. Olson for providing statistical computing assistance. **Funding:** This work was supported by the National Library of Medicine (1K99LM011014-01, R01 LM009879) and the National Institute of General Medical Sciences (R01 GM89731, R01 GM085421). **Author contributions:** A.C. and B.R. acquired the data, developed the proposed methods, and designed the study. A.C. implemented the methods and performed the experiments. A.A. and S.M. provided pharmacological expertise. A.C., B.R., A.A., and S.M. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data availability:** Requests for the 2005 and 2010 ADE data sets should be directed to Chris Madjerich (chris.madjerich@lexi.com), Product Development Manager and Clinical Liaison, Lexicomp.

Submitted 13 June 2011
Accepted 29 November 2011
Published 21 December 2011
10.1126/scitranslmed.3002774

Citation: A. Cami, A. Arnold, S. Manzi, B. Reis, Predicting adverse drug events using pharmacological network models. *Sci. Transl. Med.* **3**, 114ra127 (2011).

OPEN ACCESS Freely available online

PLOS ONE

Predicting Chemical Toxicity Effects Based on Chemical-Chemical Interactions

Lei Chen[2☯], Jing Lu[3☯], Jian Zhang[4], Kai-Rui Feng[5], Ming-Yue Zheng[3]*, Yu-Dong Cai[1]*

1 Institute of Systems Biology, Shanghai University, Shanghai, China, **2** College of Information Engineering, Shanghai Maritime University, Shanghai, China, **3** Drug Discovery and Design Center (DDDC), Shanghai Institute of Materia Medica, Shanghai, China, **4** Department of Ophthalmology, Shanghai First People's Hospital Affiliated to Shanghai Jiaotong University, Shanghai, China, **5** Simcyp Limited, Blades Enterprise Centre, Sheffield, United Kingdom

Abstract

Toxicity is a major contributor to high attrition rates of new chemical entities in drug discoveries. In this study, an order-classifier was built to predict a series of toxic effects based on data concerning chemical-chemical interactions under the assumption that interactive compounds are more likely to share similar toxicity profiles. According to their interaction confidence scores, the order from the most likely toxicity to the least was obtained for each compound. Ten test groups, each of them containing one training dataset and one test dataset, were constructed from a benchmark dataset consisting of 17,233 compounds. By a Jackknife test on each of these test groups, the 1st order prediction accuracies of the training dataset and the test dataset were all approximately 79.50%, substantially higher than the rate of 25.43% achieved by random guesses. Encouraged by the promising results, we expect that our method will become a useful tool in screening out drugs with high toxicity.

Citation: Chen L, Lu J, Zhang J, Feng K-R, Zheng M-Y, et al. (2013) Predicting Chemical Toxicity Effects Based on Chemical-Chemical Interactions. PLoS ONE 8(2): e56517. doi:10.1371/journal.pone.0056517

Editor: Gajendra P. S. Raghava, CSIR-Institute of Microbial Technology, India

Received October 16, 2012; **Accepted** January 10, 2013; **Published** February 15, 2013

Funding: This work was supported by National Basic Research Program of China (2011CB510102, 2011CB510101), National Natural Science Foundation of China (61202021, 81001399), National S&T Major Project (2012ZX09301-001-002), Hi-TECH Research and Development Program of China (2012AA020308), Innovation Program of Shanghai Municipal Education Commission (12YZ120, 12ZZ087), Shanghai Educational Development Foundation (12CG55), and Science and Technology Program of Shanghai Maritime University (No. 20120105). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: One co-author, Dr. Kai-Rui Feng is employed by a commercial company, Simcyp Limited, but he contributed to the study in his spare time. This project doesn't receive any grants or financial support from Simcyp Limited. Also, Simcyp Limited has no influence on any aspects of this research. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: myzheng@mail.shcnc.ac.cn (MYZ); cai_yud@yahoo.com.cn (YDC)

☯ These authors contributed equally to this work.

Introduction

Toxicity is a key cause of late-stage failures in drug discovery. Even some approved drugs such as Phenacetin [1] and Troglitazone [2] have been withdrawn from the market because of unexpected toxicities that were not detected during Phase III clinical trials. Thus, early toxicology data on compounds are needed to reduce R&D costs. Evaluating toxicity and assessing risks of diverse chemicals require comprehensive experimental testing against a broad spectrum of toxicity end points. These tests can cost millions of dollars, involving several thousand animals, and take many years to complete. As a result, very few chemicals have undergone the degree of testing needed to support accurate health risk assessments or meet regulatory requirements for drug approval. In recent years, the number of synthetic compounds has surged with the advance of combinatorial chemistry, and accordingly large quantities of toxicity data are urgently demanded.

Recently, particular interest has been raised to apply fast and cost-effective *in silico* toxicological models to supplement those *in vitro* and *in vivo* testing. These models require high quality toxicity data for a large set of structurally diverse drug candidates. Accelrys Toxicity is a database of toxicity information compiled from the open scientific literature [3] and containing toxicological data for approximately 0.17 million chemicals. This database is of great value for investigating the pharmacokinetic properties, metabolism and potential toxicities of compounds. Six types of toxicity data are collected in the database: (1) Acute Toxicity; (2) Mutagenicity; (3) Tumorigenicity; (4) Skin and Eye Irritation; (5) Reproductive Effects; and (6) Multiple Dose Effects. It should be noted that these categories have multiple and overlapping mechanisms of toxic action and each category represents only specific types of experiments. The combination of these experimental results may help define the overall safety profile of a compound. However, this kind of databases only provides toxicological information for recorded compounds, not for new ones. It would be valuable to accurately predict toxicities of a new compound based on the information available for recorded compounds. In order to meet the demand, there is a drive to develop quick, reliable, and non-animal-involved prediction methods, *e.g.* using structure-activity relationships (SARs) to predict drugs toxicities.

Currently, most toxicological SAR models belong to binary classifiers, which only predict compounds to be toxic or non-toxic within a single toxicity class [4,5]. It is desired to modify the strategy to predict a series of toxicity effects. In this study, we chose to build a multiclass model [6,7] to predict six categories of toxicity using the Accelrys Toxicity database instead of only one or two toxicity endpoints. However, the quadratic optimization problem

Table 1. Distribution of compounds in each category of compound toxicity.

Tag	Toxicity	Total
T_1	Acute Toxicity	12,633
T_2	Mutagenicity	6,110
T_3	Tumorigenicity	2,293
T_4	Skin and Eye Irritation	2,353
T_5	Reproductive Effects	2,501
T_6	Multiple Dose Effects	4,198
T_7	Non-toxicity	646
Total	–	30,734

doi:10.1371/journal.pone.0056517.t001

in multiclass models is difficult to solve. Thus, many previous multiclass approaches tended to decompose a multiclass problem into multiple independent binary classifications. Investigators built a set of binary classifiers, such as the model of Dietterich et al [7], each classifier distinguishing only one of the classes from the others. Although this greatly simplifies the problem, such an approach cannot provide order prediction information for the query compounds. That is, it can only predict whether the query compound has some toxicity end points, but cannot determine which is the most likely toxicity, or even the order of toxicity end points by toxicity likelihoods.

In recent years, the assessment of protein-protein interactions has been widely used to predict many attributes of proteins [8,9,10,11]. Furthermore, multiclass predictions of protein attributes have become more common [12,13,14]. These methods and their results show that interactive proteins tend to share the same functions with higher probability than do non-interactive ones. Likewise, it is reasonable to expect that interactive compounds are also more likely to share common functions as indicated by some pioneer studies [15,16]. Thus, toxicity, as part of the biological functions of compounds, should follow the same rule. Moreover, based on a previous work on the Anatomical Therapeutic Chemical (ATC) classification of drugs [16], compared to the SAR models based on physicochemical descriptors or structural alerts, a model based on chemical-chemical interactions can rank the order of the predictions more easily and yield better prediction results. In our study, we attempt to quantify chemical-chemical interactions for each pair of interactive compounds, and obtain the confidence scores of the interactions by which the toxicity end points were ordered. Briefly, compounds of seven categories including six categories of toxicity plus non-toxicity were collected. The interactive compounds of each query compound were identified utilizing STITCH (Search tool for interactions of chemicals) [17,18]. Then, the score of each class of the query compound was obtained from the confidence scores of interactions between the query compound and its interactive compounds using the toxicity profile of the interactive compounds. Finally, the prediction quality of the model was evaluated using the Jackknife test through ten test groups. Each of these was constructed from the benchmark dataset and contained one training dataset and one external test dataset. Details are described in the following sections.

Materials and Methods

Benchmark Dataset

We obtained a total of 171,266 compounds from the Accelrys Toxicity Database 2011.4 [19], which had at least one toxicity effect belonging to the following six categories: (1) Acute Toxicity; (2) Mutagenicity; (3) Tumorigenicity; (4) Skin and Eye Irritation; (5) Reproductive Effects; (6) Multiple Dose Effects. Based on compound toxicity, these compounds are allocated to the 6

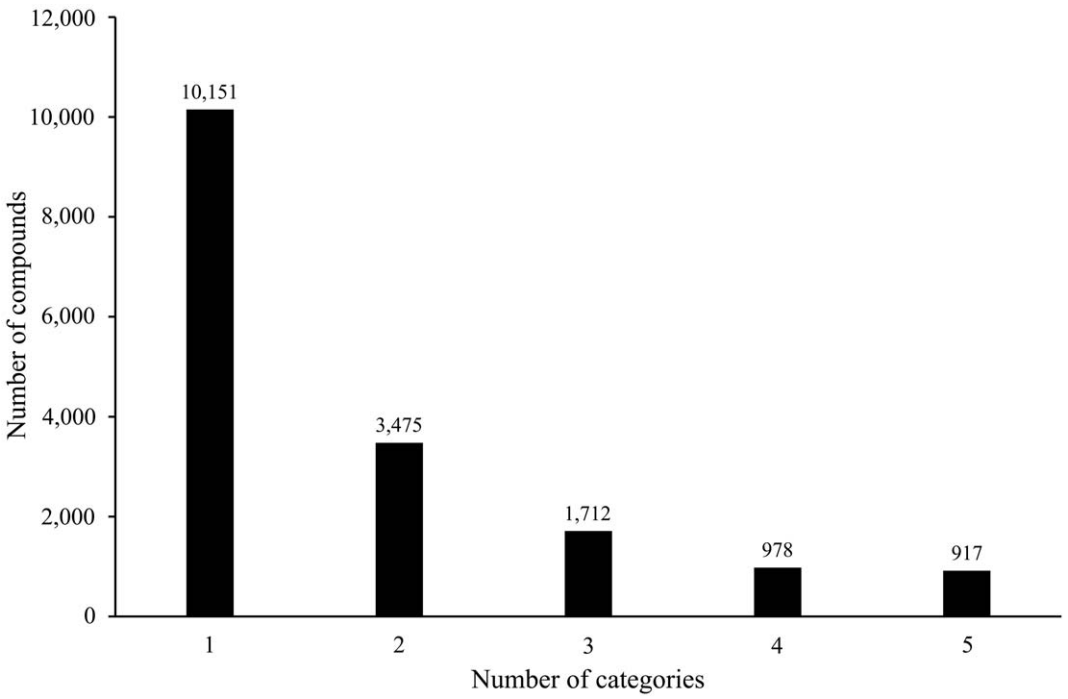


Figure 1. The number of compounds plotted against the number of categories in the benchmark dataset.
doi:10.1371/journal.pone.0056517.g001

categories, allowing multiple assignments. In addition, 2,871 "non-toxic" compounds including FDA-approved drugs from DrugBank [20] and endogenic metabolites from the Human Metabolome database (HMDB) [21] were collected and labeled as a negative class. For convenience, the 'non-toxic set' is regarded as the 7th category of compound toxicity. Due to lack of chemical-chemical interaction information in STITCH [17,18], some compounds cannot be investigated by this approach. After excluding these compounds, a benchmark dataset **S** consisting of 17,233 compounds was retrieved, of which 16,587 were toxic and 646 were non-toxic. These compounds are classified into 7 categories of compound toxicity. Shown in **Table 1** is the distribution of compounds in each category. The codes of 17,233 compounds and their toxicity information can be found in **Table S1**.

It is observed from **Table 1** that the sum of the number of compounds in all the 7 categories is much larger than the number of compounds, indicating that some compounds are allocated to more than one category of toxicity. Of the 17,233 compounds in the benchmark dataset, 10,151 compounds belong to only one category of toxicity, 3,475 compounds belong to two categories of toxicity, while others belong to 3–5 categories of toxicity and no compounds belong to more than five categories of toxicity - refer to **Figure 1** for a plot of the number of compounds against the number of categories of toxicity. Thus, prediction of compound toxicity is a multi-label classification problem. Like the case of processing proteins or compounds with multiple attributes [15,16,22], the proposed method would provide a series of candidate toxicities, ranging from the most to the least likely, instead of presenting only the most likely one.

To sufficiently evaluate the prediction method described in the following section, we constructed 10 test groups, denoted by $TG_1, TG_2, \ldots, TG_{10}$, respectively. In each test group $TG_i (1 \le i \le 10)$, there is one training dataset $\mathbf{S}_{tr}^{(i)}$ and one test dataset $\mathbf{S}_{te}^{(i)}$, *i.e.*, $TG_i = \langle \mathbf{S}_{tr}^{(i)}, \mathbf{S}_{te}^{(i)} \rangle$, where the test dataset consisted of 1,723 compounds which were randomly selected from **S**, while the training dataset contained the remaining 15,510 samples in **S**, *i.e.*, $\mathbf{S} = \mathbf{S}_{tr}^{(i)} \cup \mathbf{S}_{te}^{(i)}$ for each $1 \le i \le 10$. It is necessary to point out that, in each test group, the portion of the data in each class of the test dataset is roughly the same as that of the training dataset. Shown in **Table 2** is the distribution of compounds in training and test datasets of each test group.

Chemical-chemical Interactions

It is known that two proteins that can interact with each other are more likely to share common biological functions than non-interactive ones [8,9,10,11]. Likewise, two interactive compounds are also more likely to share similar biological functions [15,16]. Since toxicity is one of a compound's properties and functions, utilizing chemical-chemical interactions to identify compound toxicity is deemed to be feasible.

The data for chemical-chemical interactions were retrieved from STITCH (chemical_chemical.links.detailed.v3.0.tsv.gz, http://stitch.embl.de/cgi/show_download_page.) [17], a well-known database including known and predicted interactions of chemicals and proteins collected from experiments, literature or other reliable sources. In the obtained file, the interaction unit contains two compounds and five kinds of scores with titles "Similarity", "Experimental", "Database", "Textmining" and "Combined_score". The last kind of score was used here to indicate the interactivity of two compounds, *i.e.*, two compounds with "Combined_score" greater than zero were deemed interactive compounds, because the last kind of score integrates the information of the other kinds of scores. Thus, the considered interactive compounds in this study contain the following three categories: (1) those participating in the same reactions; (2) those sharing similar structures or activities and (3) those with literature associations [17]. It is known that these categories correspond to the following three facts: (I) compounds involved in the same reactions occupy the same biological pathways; (II) compounds with similar structures or activities are likely to share similar functions, thereby occupying the same pathways with high probability; (III) the co-occurrence of two compounds, as noted in many studies, indicates some direct or indirect relationships, suggesting that they have the potential to share the same pathways. On the other hand, compounds in the same biological pathways always induce similar side effects, thereby having similar toxicity effects. Accordingly, it is reasonable to suppose that interactive compounds tend to have similar toxicity effects.

The value of the "Combined_score" of two interactive compounds indicates the likelihood that they can interact, *i.e.*, two interactive compounds with high "Combined_score" can interact with high probability. Thus, this score is also termed a confidence score in this study. For two compounds c_1 and c_2, let us denote the confidence score of an interaction between them by $\mathbf{Q}(c_1,c_2)$. Specifically, if there is no interaction information between c_1 and c_2 based on the current records in STITCH, their interaction confidence score is assigned zero, *i.e.*, $\mathbf{Q}(c_1,c_2) = 0$. In this study, 323,432 interaction units, *i.e.*, 323,432 pairs of compounds with confidence scores greater than 0, were used to predict compound toxicity. The detailed information on these interaction units can be found in **Table S2**.

Prediction Method

As is mentioned in the above section, interactive compounds are more likely to have common toxicity. Accordingly, the toxicities of a query compound can be identified according to its interactive compounds.

For convenience, let T_1, T_2, …, T_7 denote the seven categories of toxicity, where T_1 denotes "Acute Toxicity", T_2 "Mutagenicity", and so forth (see column 1 and 2 of **Table 1**). Suppose that there are n compounds in the training dataset, that is c_1, c_2, …, c_n, the toxicity of a compound c_i in the training dataset is formulated as

$$T(c_i) = [t_{i,1}, t_{i,2}, \ldots, t_{i,7}] (i = 1,2,\ldots,n) \tag{1}$$

where

$$t_{i,j} = \begin{cases} 1 & \text{If } c_i \text{ has toxicity } T_j \\ 0 & \text{Otherwise} \end{cases} \tag{2}$$

Given a query compound c_q, its toxicity is predicted not only by its interactive compounds but also by the confidence scores of their interactions. The score indicating that the query compound c_q has toxicity T_j is calculated by

$$\Theta(c_q \mapsto T_j) = \sum_{i=1}^{n} \mathrm{Q}(c_i, c_q) \cdot t_{i,j} j = 1,2,3,4,5,6,7 \tag{3}$$

The high score $\Theta(c_q \mapsto T_j)$ means that there are many interactive compounds of c_q in the training dataset that have toxicity T_j or some interactions between c_q and its interactive compounds having toxicity T_j are labeled by high confidence

Table 2. Distribution of compounds in training and test datasets of each test group.

	TG_1		TG_2		TG_3		TG_4		TG_5	
Tag										
T_1	11,382	1,251	11,387	1,246	11,351	1,282	11,364	1,269	11,385	1,248
T_2	5,475	635	5,476	634	5,529	581	5,492	618	5,491	619
T_3	2,065	228	2,065	228	2,063	230	2,063	230	2,056	237
T_4	2,102	251	2,102	251	2,115	238	2,112	241	2,093	260
T_5	2,235	266	2,235	266	2,260	241	2,255	246	2,235	266
T_6	3,747	451	3,749	449	3,777	421	3,784	414	3,799	399
T_7	582	64	577	69	586	60	582	64	583	63
Total	27,588	3,146	27,591	3,143	27,681	3,053	27,652	3,082	27,642	3,092
	TG_6		TG_7		TG_8		TG_9		TG_{10}	
Tag										
T_1	11,367	1,266	11,395	1,238	11,369	1,264	11,374	1,259	11,353	1,280
T_2	5,489	621	5,500	610	5,492	618	5,497	613	5,506	604
T_3	2,075	218	2,067	226	2,070	223	2,043	250	2,070	223
T_4	2,123	230	2,125	228	2,135	218	2,102	251	2,133	220
T_5	2,244	257	2,243	258	2,236	265	2,258	243	2,234	267
T_6	3,762	436	3,750	448	3,772	426	3,777	421	3,755	443
T_7	583	63	587	59	579	67	569	77	584	62
Total	27,643	3,091	27,667	3,067	27,653	3,081	27,620	3,114	27,635	3,099

doi:10.1371/journal.pone.0056517.t002

scores. In view of this, the greater the score $\Theta(c_q \mapsto T_j)$, the more likely that the compound c_q has toxicity T_j. In particular, if $\Theta(c_q \mapsto T_j)$ for some j, it is indicated that the probability that the query c_q having the j-th category of toxicity is zero because there are no interactive compounds of c_q in the training dataset that have toxicity T_j.

Since this is a multi-label classification problem, *i.e.*, some compounds have more than one category of toxicity. A prediction method only providing the most likely toxicity is not an optimal choice. Thus, our method is valuable in that it can provide a series of candidate toxicities for a query compound, ranging from the most likely to the least likely. For example, if the results obtained from **Eq. 3** are

$$\Theta(c_q \mapsto T_3) \geq \Theta(c_q \mapsto T_1) \geq \Theta(c_q \mapsto T_6) > 0 \quad (4)$$

it can be interpreted to mean that there are three candidate toxicities for the query compound c_q, and the most likely toxicity for c_q is T_3 ("Tumorigenicity", cf. **Table 1**), followed by T_1 ("Acute Toxicity") and T_6 ("Multiple Dose Effects"). In addition, T_3 is called the 1st order prediction, T_1 the 2nd order prediction, and so forth.

Jackknife Test

The Jackknife test [16] is often used to examine the performance of various predictors, because it can always provide a unique prediction result for a given dataset. It has been widely used by investigators to evaluate their predictors [23,24,25,26,27,28,29,30,31,32,33]. During the test, each sample in the training dataset is singled out one-by-one and tested by the predictor trained by the other samples. Thus, each sample is tested exactly once.

Accuracy Measurement

The j-th order prediction accuracy is calculated by the following formula [15,16]:

$$\Gamma_j = \frac{CT_j}{N} j = 1,2,3,4,5,6,7 \quad (5)$$

where CT_j denotes the number of compounds whose j-th order prediction is one of its true toxicities, and N denotes the total number of compounds in the dataset. If a prediction method can obtain high Γ_j with small j and low Γ_j with large j, it implies that the method arranges the candidate toxicities well. Among them, the 1st order prediction accuracy is the most important indicator of good or bad performance.

Although the seven prediction accuracies can be obtained by **Eq. 5**, none of them provides the overall prediction accuracy. In view of this, we employ another measurement that calculates the proportion of true toxicities of the first m predictions. It can be calculated as follows [16]:

$$\Delta_m = \frac{\sum_{i=1}^{N} S_{i,m}}{\sum_{i=1}^{N} N_i} \quad (6)$$

where $S_{i,m}$ represents the number of the correct predictions of the i-th compound among its first m predictions, and N_i represents the number of toxicities that the i-th compound has. Since different compounds may have different numbers of toxicities, the parameter m in **Eq. 6** is usually taken as the smallest integer no less than the average number of toxicities in the dataset, which can

Table 3. Prediction accuracies obtained by the method as applied to training and test datasets of each test group.

	TG_1		TG_2		TG_3		TG_4		TG_5	
Prediction Order										
1	79.40%	79.69%	79.45%	79.28%	79.23%	80.62%	79.28%	79.45%	79.30%	79.34%
2	37.16%	38.42%	37.14%	38.24%	37.54%	37.20%	37.17%	38.31%	37.40%	36.16%
3	22.18%	23.16%	22.20%	22.87%	22.32%	21.65%	22.29%	22.63%	22.53%	22.87%
4	15.45%	16.66%	15.49%	16.77%	16.35%	14.86%	15.46%	16.13%	15.41%	15.55%
5	11.06%	11.61%	11.04%	11.49%	11.00%	10.85%	10.88%	10.16%	10.95%	11.20%
6	6.92%	7.25%	6.84%	7.89%	7.23%	5.86%	6.99%	6.56%	6.85%	7.84%
7	1.21%	1.33%	1.22%	1.04%	1.27%	1.51%	1.39%	1.45%	1.26%	1.68%
	TG_6		TG_7		TG_8		TG_9		TG_{10}	
Prediction Order	$S_{tr}^{(6)}$	$S_{te}^{(6)}$	$S_{tr}^{(7)}$	$S_{te}^{(7)}$	$S_{tr}^{(8)}$	$S_{te}^{(8)}$	$S_{tr}^{(9)}$	$S_{te}^{(9)}$	$S_{tr}^{(10)}$	$S_{te}^{(10)}$
1	79.57%	80.15%	79.36%	79.98%	79.45%	79.05%	79.52%	79.80%	79.46%	79.34%
2	37.11%	37.72%	37.57%	36.10%	37.21%	38.65%	37.32%	35.98%	37.44%	37.20%
3	22.57%	22.29%	22.30%	23.39%	22.23%	24.03%	22.46%	23.33%	22.42%	22.93%
4	15.31%	15.90%	15.36%	15.55%	15.52%	14.74%	15.40%	16.25%	15.36%	16.37%
5	10.93%	10.45%	10.95%	11.55%	11.08%	10.10%	10.74%	11.55%	10.87%	10.74%
6	7.00%	6.56%	7.00%	6.62%	7.16%	5.86%	6.76%	7.78%	6.97%	7.25%
7	1.25%	1.57%	1.32%	0.99%	1.32%	1.45%	1.27%	1.57%	1.30%	1.33%

doi:10.1371/journal.pone.0056517.t003

be computed by

$$M = \frac{\sum_{i=1}^{N} N_i}{N} \tag{7}$$

where $m = \lceil M \rceil$. Obviously, a larger Δ_m implies better prediction performance by the method for the identification of compound toxicity.

Table 4. Proportions of true toxicities covered by the first two predictions for training and test datasets of each test group.

Test group	Training dataset	Test dataset
TG_1	65.52%	64.69%
TG_2	65.54%	64.52%
TG_3	65.43%	66.49%
TG_4	65.32%	65.83%
TG_5	65.48%	64.36%
TG_6	65.46%	65.71%
TG_7	65.55%	65.21%
TG_8	65.43%	65.82%
TG_9	65.61%	64.07%
TG_{10}	65.61%	64.79%

doi:10.1371/journal.pone.0056517.t004

Results

As described in the Section "Benchmark dataset", 10 test groups were constructed to evaluate the method described in Section "Prediction method". In each test group, there were one training dataset consisting of 15,510 compounds and one test dataset containing 1,723 compounds. The predicted results for each test group obtained by the proposed method are as follows.

Performance of the Method on the Training Dataset

For the 15,510 compounds in each training dataset $S_{tr}^{(i)}(1 \le i \le 10)$, we conducted the prediction and evaluated its performance by the Jackknife test. Listed in the column with title $S_{tr}^{(i)}$ of **Table 3** are seven prediction accuracies, calculated by **Eq. 5**, for training dataset $S_{tr}^{(i)}$, from which we can see that the 1st order prediction accuracies were all around 79.50%, where the maximum was 79.57%, while the minimum was 79.23%; the 2nd order ones were all around 37.30%. It is indicated that the proposed method is very stable. It is also observed from the corresponding columns of **Table 3** that the accuracies followed a descending trend when increasing the order number, indicating that the method sorted the candidate toxicities quite well for the compounds in each training dataset $S_{tr}^{(i)}(1 \le i \le 10)$. The average

Figure 2. The structures of the alkyl N-nitroso group and the primary aromatic amine group.
doi:10.1371/journal.pone.0056517.g002

Table 5. Details of Tasosartan's interactive compounds in the training dataset.

Compound ID	Tag of toxicity class	Its interactive compound ID	Tag of toxicity class	Confidence score
CID000060919	T_7	CID000003749	T_7	679
CID000060919	T_7	CID000002541	T_7	670
CID000060919	T_7	CID000060921	T_7	669
CID000060919	T_7	CID000003961	T_7	667
CID000060919	T_7	CID000060846	T_7	658
CID000060919	T_7	CID000065999	T_1, T_6	643
CID000060919	T_7	CID000054738	T_1, T_2	172

doi:10.1371/journal.pone.0056517.t005

numbers of toxicities for compounds in each training dataset $S_{tr}^{(i)}$ were about 1.78 according to **Eq. 7**, *i.e.*, $M = 1.78$. It is noteworthy that if one predicts compound toxicity by random guesses, the average success rate would be only 25.43% (1.78/7), which is much lower than each of the 1st order prediction accuracies by our method. To evaluate the prediction accuracy by the method more thoroughly, **Eq. 6** was calculated by taking $m = 2$, *i.e.*, we considered the first two predictions for each compound in $S_{tr}^{(i)}(1 \le i \le 10)$ to see the proportions of true toxicities covered by these predictions. These proportions are shown in column 2 of **Table 4**, from which we can see that they were all about 65.50%, where the maximum was 65.61% while the minimum was 65.32%. Thus, it is indicated once again that our method is reliable.

Performance of the Method on the Test Dataset

For the 1,723 compounds in each test dataset $S_{te}^{(i)}(1 \le i \le 10)$, the toxicities of these compounds were predicted by the proposed method described in Section "Prediction method" based on the compounds in the training dataset $S_{tr}^{(i)}$. After processing by **Eq. 5**, seven prediction accuracies for each test dataset $S_{te}^{(i)}$ were obtained and were listed in the column with title $S_{te}^{(i)}$ of **Table 3**. It is observed that the 1st order prediction accuracies were all about 79.50%. Similar to the seven prediction accuracies for each training dataset $S_{tr}^{(i)}$, those of test dataset $S_{te}^{(i)}$ also followed a descending trend with the increase of the order number, implying that our method also arranged the candidate toxicities of samples in each test dataset quite well. According to **Eq. 7**, the average numbers of toxicities for the compounds in each test dataset were about 1.80. Thus, we still considered the first two predictions of each sample in $S_{te}^{(i)}(1 \le i \le 10)$ to calculate the proportions of true toxicities covered by these predictions, *i.e.*, computing **Eq. 6** by taking $m = 2$. Listed in column 3 of **Table 4** are ten proportions for ten test datasets, each yielding a probability of approximately 65%.

Table 6. The details of common compounds belonging to two categories.

Tag of toxicity class	T_1	T_2	T_3	T_4	T_5	T_6
T_1	12,633[a]	3,483 (22.8%)[b]	1,485 (11.0%)	2,027 (15.6%)	2,075 (15.9%)	3,446 (25.7%)
T_2		6110	1,720 (25.7%)	1,213 (16.7%)	1,336 (18.4%)	1,723 (20.1%)
T_3			2293	570 (14.0%)	753 (18.6%)	781 (13.7%)
T_4				2353	731 (17.7%)	897 (15.9%)
T_5					2501	1,409 (26.6%)
T_6						4,198

[a]The number of common compounds belonging to two categories.
[b]The number in parenthesis means the ratio of the number of common compounds to the number of non-overlapping compounds of the two categories.
doi:10.1371/journal.pone.0056517.t006

Discussion

Understanding of the Toxicity Prediction Results

It is observed from **Table 3** that the performance of the method on ten test groups is similar. Thus, the first test group (*i.e.*, TG_1) is used as an example to show how to interpret the toxicity predicting results in detail.

Our multiclass model achieved a quite promising performance using the chemical-chemical interactions data on test group TG_1 (see **Table 3** for details). For example, the compound 4-(N-methyl-N-nitrosamino)-1-(3-pyridyl)-1-butanone (CID000047289, NNK) shows positive results for five toxicity endpoints: T_1, T_2, T_3, T_5, and T_6. Our model accurately predicted these five kinds of endpoints, and provided the order predictions as $T_3 > T_2 > T_1 > T_6 > T_5 > T_4 > T_7$. The 7th label representing 'non-toxic' was ranked as the last, suggesting that this compound is very likely to have toxic effects. As stated in the Section "Chemical-chemical interactions", the interactive compounds derived from STITCH tend to have the same toxicity categories. 4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanol (CID000104856, NNAL), an interactive compound of NNK, has toxicities T_2 and T_3, which are also shared by NNK. The alkyl N-nitroso group (see **Figure 2**) of these two compounds associates with the formation of DNA adducts, and induces lung cancer in laboratory animals [34,35,36]. Another example is trimethoprim (CID000005578), which is positive for five toxicity endpoints: T_1, T_2, T_4, T_5, and T_6. The prediction order of our model was $T_1 > T_6 > T_2 > T_5 > T_4 > T_3 > T_7$. This compound was considered to be a carcinogen according to chemical-chemical interactions, but the Accelrys Toxicity database [19] labeled this compound only as a mutagen. However, it is reasonable to assume this compound as a carcinogen because it has a genotoxic toxicophore-aromatic amine (see **Figure 2**) [5,37,38]. Typically, mutation is one of the first steps in the development of cancer [39].

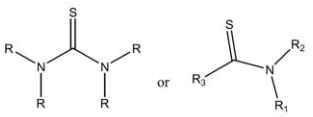
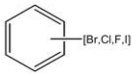
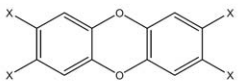
Nongeneric SAs	Examples
 thiocarbonyl	diphenylthiohydantoin (CID000854150) rubeanic acid (CID002777982)
 halogenated benzene	4-chlorothiophenol (CID000007815) 4-bromophenol (CID000007808)
halogenated PAH	4-hydroxy-2',4',6'-trichlorobiphenyl (CID000105036) 2,2',5,5'-tetrachlorobenzidine (CID000027465)
 halogenated dibenzodioxins	1,2,3,4,6,7,8-heptachlorodibenzodioxin (CID000037270) 2,3,7,8-tetrabromodibenzo-4-dioxin (CID000039729)

Figure 3. Nongeneric SAs (Benigni) and some carcinogens matching these SAs.
doi:10.1371/journal.pone.0056517.g003

Tasosartan (CID000060919) is an angiotensin II (AngII) receptor blocker [40], which is labeled as a relatively "non-toxic" compound in the dataset. Using our model, the order prediction of this compound was $T_7 > T_1 > T_6 > T_2$. The 1st order prediction is "non-toxic", consistent with the experimental data available. Among seven interactive compounds in the training dataset retrieved from STITCH (see **Table 5**), the top five interactive compounds are "non-toxic", and their confidence scores are relatively high. However, the latter two interactive compounds are toxic, so tasosartan is predicted to have some toxicity effects in our model. However, the possibility of its possessing these toxicities is less than that of its not possessing toxicity (*i.e.*, "non-toxic").

The predictions for NNK, trimethoprim, and tasosartan and the prediction accuracies of the method indicate that interactive compounds can share common toxicity with high probability, which assessment conforms to the results of predicting other attributes of compounds [15,16]. The confidence scores of chemical-chemical interactions contribute significantly to the prediction of compound toxicity. As shown in **Table 5**, the interactive compounds of tasosartan with high confidence scores dominantly have the same toxicity as tasosartan. On the other hand, the predicted results for NNK, trimethoprim, and tasosartan reflect a limitation of our model: the judgment of "toxic" or "non-toxic" is based on a collective set of compounds with interactive information. However, some compounds with low confidence scores exist and they may contribute to the input of promiscuous interaction information to the final classification model. To address this issue, a future endeavor should introduce a threshold to the interaction confidence score and exclude "noisy" information to obtain a more accurate prediction.

Moreover, many more compounds are without chemical-chemical interactions in the original Accelrys Toxicity database. It is expected that the problem of predicting compound toxicity can be solved more favorably by the method as increasing amounts of chemical-chemical interaction information become available.

Analysis of the Relationship between Different Chemical Toxicity Effects

In the Accelrys Toxicity Database, there are 3,607 compounds with more than two types of toxicity effects and 3,475 compounds with exact two effects (refer to **Figure 1**). We analyzed the number of common compounds belonging to two categories, and the ratio of the number of common compounds to the number of non-overlapping compounds of the two categories (see **Table 6**). It can be found that the intersection of T_5 ("Reproductive Effects", cf. **Table 1**) and T_6 ("Multiple Dose Effects") is the largest, sharing 26.6% of common compounds. The overlapping compounds suggest that there may be a causal relationship between the two categories. Specifically, the reproductive effects may cause multiple dose effects, *i.e.*, reproductive toxicities may be cumulative, and hence be regarded as showing multiple dose effects in the meantime. The followed instances of correspondence between two categories are T_2 ("Mutagenicity") vs. T_3 ("Tumorigenicity") and T_1 ("Acute Toxicity") vs. T_6 ("Multiple Dose Effects"). Since, in many cases, mutation is one of the first steps in the development of cancer [39], we took T_2 ("Mutagenicity") vs. T_3 ("Tumorigenicity") as an example to study the relationship between the two toxic categories.

From the viewpoint of mechanism of action, carcinogens can be classified into genotoxic or epigenetic carcinogens. Genotoxic carcinogens can bind covalently to DNA, and many known mutagens belong to this category. In the dataset, there are 1,720 common compounds with simultaneous toxicity T_2 ("Mutagenicity") and T_3 ("Tumorigenicity"). The Structural alerts (SAs) provided by Benigni [37], which are molecular functional groups associated with a specific toxicity end point [38], were used here to gain insights into the correspondence of the two toxic effects. As summarized in **Table S3**, we illustrated a few examples for each of the matched SAs.

As previously mentioned, not all of the mutagens are carcinogens. For example, α,β-unsaturated carbonyl compounds can interact with DNA by Michael addition, then lead to mutagenic and carcinogenic responses [37], *e.g.* acrylamide (CID000006579) and 2-butenal (CID000447466). However, if an

α,β-unsaturated carbonyl compound has conformational constraints or alkyl groups at the site of nucleophilic attack, the compound would be prone to reaction via Schiff base formation [41]. This change may only generate the DNA-adducts, but not undergo the following carcinogenic process [37]. This means that this kind of compound has no carcinogenicity, *e.g.* (E)-2-methyl-2-butenal (CID005321950) and 2-propylacrolein (CID000070609).

Epigenetic carcinogens do not usually bind directly to DNA, but have a large variety of different and specific mechanisms, and behave negatively in the standard mutagenicity assay [42]. Thus, some compounds that can match nongeneric SAs [37] are only carcinogens, not mutagens (see **Figure 3**).

Conclusions

In this study, a multi-classifier for six toxicity effects was built based on 17,233 compounds with their experimental toxicity information available and 323,432 pairs of mapped chemical-chemical interaction information extracted from the STITCH database. A new chemical entity can have multiple toxicity effects, so a multiclass toxicity prediction tool may prove to be practically more valuable to chemists than a traditional binary classification model. It can provide a better toxicity profile for a compound rather than merely indicating whether the compound has a specific toxic action or potential. The outstanding performance of our approach suggests that the multi-classification scheme is feasible and effective for *in silico* chemical toxicity prediction.

Supporting Information

Table S1 List of 17,233 compounds investigated in this study and their toxicity information.
(PDF)

Table S2 List of 323,432 interaction units used to predict compound toxicity in this study.
(PDF)

Table S3 List of SAs (Benigni) and examples matching SAs in our dataset.
(PDF)

Author Contributions

Conceived and designed the experiments: LC JZ MYZ YDC. Performed the experiments: LC JL KRF. Analyzed the data: JL JZ MYZ. Contributed reagents/materials/analysis tools: LC JL KRF MYZ YDC. Wrote the paper: LC JL KRF.

References

1. Dubach UC, Rosner B, Sturmer T (1991) An epidemiologic study of abuse of analgesic drugs. Effects of phenacetin and salicylate on mortality and cardiovascular morbidity (1968 to 1987). N Engl J Med 324: 155–160.
2. "AstraZeneca Decides to Withdraw Exanta" (2006) Available: http://www.astrazeneca.com/Media/Press-releases/Article/20060214-AstraZeneca-Decides-to-Withdraw-Exanta.Accessed 2012 Sep 2.
3. Wang WB, Zhao YP, Cong L, Jing H, Liao Q, et al. (2011) Clinical characters of gastrointestinal lesions in intestinal Behcet's disease. Chin Med Sci J 26: 168–171.
4. Zheng M, Liu Z, Xue C, Zhu W, Chen K, et al. (2006) Mutagenic probability estimation of chemical compounds by a novel molecular electrophilicity vector and support vector machine. Bioinformatics 22: 2099–2106.
5. Wang Y, Lu J, Wang F, Shen Q, Zheng M, et al. (2012) Estimation of carcinogenicity using molecular fragments tree. J Chem Inf Model 52: 1994–2003.
6. Crammer K, Singer Y (2001) On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research 2: 265–292.
7. Dietterich TG, Bakiri G (1995) Solving multiclass learning problems via error-correcting output codes. Journal of Artificial Intelligence Research 2: 263–286.
8. Sharan R, Ulitsky I, Shamir R (2007) Network-based prediction of protein function. Mol Syst Biol 3: 88.
9. Bogdanov P, Singh AK (2010) Molecular function prediction using neighborhood features. IEEE/ACM Trans Comput Biol Bioinform 7: 208–217.
10. Kourmpetis YA, van Dijk AD, Bink MC, van Ham RC, ter Braak CJ (2010) Bayesian Markov Random Field analysis for protein function prediction based on network data. PLoS One 5: e9293.
11. Ng KL, Ciou JS, Huang CH (2010) Prediction of protein functions based on function-function correlation relations. Comput Biol Med 40: 300–305.
12. Hu L, Huang T, Liu XJ, Cai YD (2011) Predicting protein phenotypes based on protein-protein interaction network. PLoS One 6: e17668.
13. Hu L, Huang T, Shi X, Lu WC, Cai YD, et al. (2011) Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties. PLoS One 6: e14556.
14. Gao P, Wang QP, Chen L, Huang T (2012) Prediction of Human Genes Regulatory Functions Based on Proteinprotein Interaction Network. Protein and Peptide Letters 19: 910–916.
15. Hu LL, Chen C, Huang T, Cai YD, Chou KC (2011) Predicting Biological Functions of Compounds Based on Chemical-Chemical Interactions. PLoS ONE 6: e29491.
16. Chen L, Zeng WM, Cai YD, Feng KY, Chou KC (2012) Predicting Anatomical Therapeutic Chemical (ATC) Classification of Drugs by Integrating Chemical-Chemical Interactions and Similarities. PLoS ONE 7: e35254.
17. Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P (2008) STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 36: D684–688.
18. Kuhn M, Szklarczyk D, Franceschini A, Campillos M, von Mering C, et al. (2010) STITCH 2: an interaction network database for small molecules and proteins. Nucleic Acids Res 38: D552–556.
19. Accelrys Toxicity Database 2011.4. Accelrys Software Inc.: San Diego, CA.
20. DrugBank. Available: http://www.drugbank.ca/downloads. Accessed 2012 Sep 2.
21. HMDB. Available: http://www.hmdb.ca/downloads. Accessed 2012 Sep 2.
22. Du P, Li T, Wang X (2011) Recent progress in predicting protein sub-subcellular locations. Expert Review of Proteomics 8: 391–404.
23. Cai YD, Lu L, Chen L, He JF (2010) Predicting subcellular location of proteins using integrated-algorithm method. Molecular Diversity 14: 551–558.
24. Shao X, Tian Y, Wu L, Wang Y, Jing L, et al. (2009) Predicting DNA-and RNA-binding proteins from sequences with kernel methods. Journal of Theoretical Biology 258: 289–293.
25. Zeng Y, Guo Y, Xiao R, Yang L, Yu L, et al. (2009) Using the augmented Chou's pseudo amino acid composition for predicting protein submitochondria locations based on auto covariance approach. Journal of Theoretical Biology 259: 366–372.
26. Chen L, Cai YD, Shi XH, Huang T (2012) Analysis of Metabolic Pathway Using Hybrid Properties. Protein and Peptide Letters 19: 99–107.
27. Esmaeili M, Mohabatkar H, Mohsenzadeh S (2010) Using the concept of Chou's pseudo amino acid composition for risk type prediction of human papillomaviruses. Journal of theoretical biology 263: 203–209.
28. Georgiou D, Karakasidis T, Nieto J, Torres A (2009) Use of fuzzy clustering technique and matrices to classify amino acids and its impact to Chou's pseudo amino acid composition. Journal of theoretical biology 257: 17–26.
29. Li BQ, Hu LL, Chen L, Feng KY, Cai YD, et al. (2012) Prediction of Protein Domain with mRMR Feature Selection and Analysis. PLoS ONE 7: e39308.
30. Jin L, Fang W, Tang H (2003) Prediction of protein structural classes by a new measure of information discrepancy. Computational Biology and Chemistry 27: 373–380.
31. Ivanciuc O (2008) Weka machine learning for predicting the phospholipidosis inducing potential. Current Topics in Medicinal Chemistry 8: 1691–1709.
32. Ravetti MG, Moscato P (2008) Identification of a 5-protein biomarker molecular signature for predicting Alzheimer's disease. PLoS ONE 3: e3111.
33. Sun XD, Huang RB (2006) Prediction of protein structural classes using support vector machines. Amino acids 30: 469–475.
34. Yuan JM, Koh WP, Murphy SE, Fan Y, Wang R, et al. (2009) Urinary levels of tobacco-specific nitrosamine metabolites in relation to lung cancer development in two prospective cohorts of cigarette smokers. Cancer Res 69: 2990–2995.
35. Kitiporn P, Jan-Phillip M, Marcus O, Fabian S, Victor S, et al. (2008) Machine learning based analyses on metabolic networks supports high-throughput knockout screens. BMC Systems Biology 2: 67.
36. Church TR, Anderson KE, Caporaso NE, Geisser MS, Le CT, et al. (2009) A prospectively measured serum biomarker for a tobacco-specific carcinogen and lung cancer in smokers. Cancer Epidemiol Biomarkers Prev 18: 260–266.
37. Benigni R, Bossa C (2011) Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. Chem Rev 111: 2507–2536.
38. Benigni R, Bossa C (2008) Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutat Res 659: 248–261.
39. Arcos JC, Argus MF, editors (1995) Multifactor interaction network of carcinogenesis - a "tour guide". Boston: Birkhauser. 1–20 p.

40. DrugBank. Available: http://www.drugbank.ca/drugs/DB01349. Accessed 2012 Sep 12.
41. Patlewicz GY, Wright ZM, Basketter DA, Pease CK, Lepoittevin JP, et al. (2002) Structure-activity relationships for selected fragrance allergens. Contact Dermatitis 47: 219–226.
42. Woo YT (2003) Mechanisms of action of chemical carcinogens, and their role in structure–activity relationships (SAR) analysis and risk assessment. In: Benigni R, editor. Quantitative Structure–Activity Relationship (QSAR) Models of Mutagens and Carcinogens. Boca Raton: CRC Press. 41–80.

DRUG DEVELOPMENT

Predicting toxicity

DOI:
10.1038/nrd2470

More efficient compound safety screening methods are much-needed to help reduce the persistently high attrition rates in clinical development. Now, writing in the *Journal of Proteome Research*, Nicholson and colleagues present the novel approach taken by the Consortium on Metabonomic Toxicology (COMET), which provides the largest validation so far of the potential of metabolic profiling to predict drug toxicity.

Their metabonomics-based method involved profiling compound-induced perturbations in urinary metabolites using ^{1}H nuclear magnetic resonance spectroscopy (NMR). Their study included 80 compounds that were selected to cover a diverse range of structures and toxicities, with an emphasis on liver and kidney toxins as these are the major organs involved in toxicity.

Urine samples were collected from male rats both prior to and at various time points over 7 days following treatment with a single dose of compound. This culminated in 6,260 control and 6,675 treated urine samples from 1,652 rats. Histopathology and clinical chemistry evaluations were carried out at 48 and 168 hours post-dose, respectively, to monitor toxicity.

To assess the effect of compounds on urinary metabolites, the authors first built a multivariate model of normal urine based on pre-processed ^{1}H NMR spectra of the samples. Classification of samples from dosed animals as normal or abnormal using this model revealed a high correspondence between toxicity and abnormal metabolic profiles, with 67 out of the 80 treatments showing agreement as to the presence or absence of an effect. Compared with the normal model, 62 treatments exerted an effect and these were used for subsequent studies.

Next, Nicholson *et al.* set out to determine whether urinary metabolite analysis could be used to detect specific organ toxicity. To do this, they used a density estimation method — Classification of Unknowns by Density Superposition (CLOUDS). This combines NMR data obtained from all animals across all time points within the studies for a particular treatment, which can then be compared as a single unit with the signatures of other treatments.

Using the CLOUDS method in blind tests the authors could correctly identify the target organ of the liver toxin azathioprine and the kidney toxin maleic acid, even at sub-toxic levels. Assessment of the system across all 62 treatments showed that it had a sensitivity — the proportion of all treatments affecting a given organ that are classified to that organ — to liver and kidney toxins of 67% and 41%, respectively. The corresponding specificities — the proportion of all treatments predicted to affect a given organ that truly affect that organ — were 77% and 100%, respectively.

These promising results indicate that this metabonomics-based method provides a non-invasive, sensitive, rapid and cost-effective approach for preclinical toxicology, and it is currently in use by several of the COMET pharmaceutical partners. Such an approach could also have potential in studying drug efficacy in preclinical studies and clinical trials.

Sarah Crunkhorn

ORIGINAL RESEARCH PAPER Ebbels, T. *et al.* Prediction and classification of drug toxicity using probabilistic modeling of temporal metabolic data: the consortium on metabonomic toxicology screening approach. *J. Prot. Res.* **6**, 4407–4422 (2007)
FURTHER READING Lindon, J. *et al.* The Consortium for Metabonomic Toxicology (COMET): aims, activities and achievements. *Pharmacogenomics* **6**, 691–699 (2005)

The International Serious Adverse Events Consortium

Arthur L. Holden, Jorge L. Contreras, Sally John and Matthew R. Nelson

The International Serious Adverse Events Consortium is generating novel insights into the genetics and biology of drug-induced serious adverse events, and thereby improving pharmaceutical product development and decision-making.

Arthur L. Holden is Founder, CEO and Chairman of the iSAEC, 8770 West Bryn Mawr Avenue, Suite 1300, Chicago, Illinois 60631, USA.
Jorge L. Contreras is Legal Counsel to the iSAEC and at the S.J. Quinney College of Law, and Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA.
Sally John is Co-Chairperson of the iSAEC Scientific Management Committee and at Pfizer, New London, Connecticut 06320, USA.
Matthew R. Nelson is Co-Chairperson of the iSAEC Scientific Management Committee and at GlaxoSmithKline, Research Triangle Park, North Carolina 27709–3398, USA.
Correspondence to A.L.H.
e-mail: aholden@earthlink.net
doi:10.1038/nrd4441

The impetus for the International Serious Adverse Events Consortium (iSAEC) arose from a series of interviews in 2006 with senior research and development leaders of major pharmaceutical companies, exploring how to build on the success of the SNP Consortium[1] to identify additional, high-value genomic research areas in which to apply this highly effective cross-industry collaborative model. The interviewees assigned the highest priority to exploring the genetic basis of drug-induced, rare serious adverse events (SAEs). In May 2006, with staff at the US Food and Drug Administration (FDA), we conceptualized the structure for a private, international research consortium to explore the genetic contribution to drug-induced SAEs. It was felt the opportunities for applying genomic technologies to better understand this vital aspect of drug safety would benefit both drug development and regulatory oversight. Equally significant were the complexity, logistics, management, risks, and cost associated with such a research initiative. No single institution possessed the resources, sufficient well-phenotyped cases, genomics expertise and international breadth to execute such a research endeavour alone. The stage was set for the development and launch of the iSAEC.

Scientific focus and organizational structure

The iSAEC is a pharmaceutical-industry-led and FDA-supported international research consortium, focused on identifying and validating DNA variants predictive of the risk of drug-induced SAEs. It was launched in 2007 with the scientific and financial support of six funding members (Abbott, GlaxoSmithKline, Johnson & Johnson, Pfizer, Roche and Sanofi-Aventis). Additional dues-paying members were added (Novartis, Takeda, Daiichi Sankyo, and The Wellcome Trust) as the consortium completed its Phase 1 research programme (focused on the genetics of drug-induced liver injury (DILI) and serious skin injury (DISI)). A separate call for funding and membership roster was developed for the Phase 2 research programme, which included ten dues-paying members (Abbott, GlaxoSmithKline, Pfizer, Takeda, Daiichi Sankyo, Novartis, Merck, Amgen, AstraZeneca and the Wellcome Trust), as well as three associate members that made in-kind, non-cash contributions to the research effort (Cerner, Clinical Data and Catholic Health Initiatives). The FDA has participated from the outset as an observer, advisor and research collaborator, but without formal membership status.

Since 2007, the iSAEC has collaborated with over 200 leading academic centres and scientists globally to:

- standardize and publish phenotype definitions for the major drug-induced SAEs (liver, skin, heart and renal injury);
- build diverse, well-phenotyped clinical cohorts and sample repositories for many of the major SAEs;
- apply optimal genomic and computational methods (including imputation) for effective genome-wide single nucleotide polymorphism (SNP) genotyping and exome sequencing;
- ensure timely public availability of scientific results/associated data (within 12 months after genotyping, regardless of publication timing) to the scientific community at no cost[2] (see Further information); and
- ensure the open use of all iSAEC data, unencumbered by intellectual property constraints[3].

The iSAEC's organization is virtual and composed of multiple collaborative teams, staffed by member volunteers and research collaborators, and under the direction of the iSAEC's CEO/Chairman. The iSAEC is governed by a board of directors (BOD) that consists of one director from each sponsoring member and the CEO, *ex officio*, and makes its decisions using a 'majority rules' model. The board delegates the oversight and management of the consortium's research agenda to the scientific management committee (SMC), which has representatives from each member company as well as scientific and clinical experts from many of its major collaborations. The SMC is supported by the Data Analysis and Coordination Center (DACC) at Columbia University as well as a network of genotyping and sequencing partners. The DACC coordinates the aggregation, quality control, analysis and release of all research data; prior to public data release, no

consortium member or collaborator may use the data for any purpose other than the advancement of the consortium's research (that is, there is no preferential access; see Further information for details of the data release policy).

Current status and scientific output so far

Over the past 7 years, the iSAEC has developed novel, international clinical networks to aggregate well-phenotyped case collections associated with specific SAEs and causal drugs. Specifically, we have aggregated subjects with DILI, DISI, drug-induced hypersensitivity syndrome (DIHSS), drug-induced renal injury (DIRI), drug-induced Torsades de pointes/prolonged QT effects (DITdP), inflammatory bowel disease (IBD) therapy-related SAEs such as pancreatitis and leukopenia, excessive weight gain (EWG) associated with class 2 antipsychotics, and osteonecrosis of the jaw (ONJ). Case enrolment has been completed for all SAEs, with the exception of DIRI and those related to IBD (see Supplementary information S1 (table)). By the end of 2015, the consortium expects to have aggregated close to 7,500 SAE cases spanning these phenotypes. The majority of this collection will be Caucasian, but it will contain important African, Indian and Chinese cohorts. The scale, depth, quality, and diversity of this recruitment effort are unprecedented in the history of drug safety research.

The iSAEC has or will conduct genome-wide genotyping of all collected subjects. In Phase 1, initial genome-wide association studies were conducted for DILI, DISI and DITdP, leading to several novel findings and key insights into the primary immune-related mechanisms underlying many of these SAEs (see Supplementary information S2 (box) for a list of publications). Following the success of the first phase, the BOD approved a plan to increase the existing DILI and DISI case collections, expand into DIHSS, DIRI, EWG, ONJ and IBD-related SAEs, expand investigations for selected SAEs into non-European populations, and explore the role of rare variants in SAEs with pilot exome sequencing studies for co-amoxiclav-induced liver injury, clozapine-induced agranulocytosis and DITdP.

To date, the iSAEC has completed 18 public releases of anonymized subject-level clinical and genotyping data, associated with 3,623 of its cases and controls. A total of 135 researchers and institutions have applied for and been granted access to the iSAEC database (see Further information). Through this open access policy, we hope to stimulate further analysis that will yield additional scientific insights and publications as collections and genetic analysis methods evolve[2].

The iSAEC is helping to set the precedent for genetic analysis of drug-induced SAEs and beginning to broaden the scientific understanding of these highly personalized reactions to otherwise safe and effective drugs. Through our research, we have demonstrated that the primary genetic contribution to SAE risk is through human leukocyte antigen (HLA) variation and the adaptive immune response, and that the variants with clinically meaningful effects can be detected in relatively small sample sizes (<50 cases in several instances). This bodes well for the feasibility of applying genomic methods in the future when an immunologically mediated toxicity is suspected. In those studies where we have performed sequencing analysis, our quest to identify rare variants (that is, <1% of the population) with a large SAE influence has, to date, been unfruitful. We remain uncertain as to the effects such rare genetic variants may have on SAEs. To date, most of our findings are drug-specific versus across multiple drugs, which may be expected given the important role for the major histocompatibility complex genomic region in the pathology of immunologically mediated SAEs and the very specific relationships observed between HLA alleles and clinical disease (for example, HLA-B*27 in ankylosing spondylitis and HLA-C*06 in psoriasis). Finally, there are a number of HLA alleles that are associated with different SAEs and for different drugs, including HLA-B*57:01, HLA-DRB1*07:01, and HLA-DRB1*15:01, that may provide important insights into the underlying biology of SAEs and offer strategies to predict or mitigate future SAEs.

Lessons learned and conclusions

Lessons learned in developing the iSAEC include:

- a clear, unifying, highly important mission is a must from the outset;
- to maximize membership and ease of formation, ensure the proposed effort is precompetitive and in the public good;
- develop the operating plan and uniform membership requirements with the potential funding members;
- have a high-quality, phased scientific/operating plan before recruiting funding members;
- establish dedicated, high-quality management early;
- develop funding requirements early, and work with the potential members on trade-offs to produce an affordable and effectively phased consortium;
- organize a board and well-defined committees with high-quality, dedicated leaders;
- outsource to the best external advisors/investigators via performance-based contracts;
- exceed expectations;
- make it fun and say "thank you" in meaningful ways;
- know when to terminate the consortium — begin with the end in mind!

Drug-related biomedical research options are exploding in number, complexity, risk and cost. To address the challenges, all stakeholders must work together to develop new collaborative research frameworks and diversified funding models that enhance financial leverage and research productivity. The iSAEC serves as an excellent example of such innovation.

1. Holden, A. L. The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques* **32**, S22–S24 (2002).
2. Contreras, J. L., Floratos, A. & Holden, A.L. The International Serious Adverse Events Consortium's data sharing model. *Nature Biotech.* **31**, 17–19 (2012).
3. Contreras, J. L. Bermuda's legacy: policy, patents, and the design of the genome commons. *Minn. J. Law Sci. Technol.* **12**, 61–125 (2011).

Competing interests statement
The authors declare no competing interests.

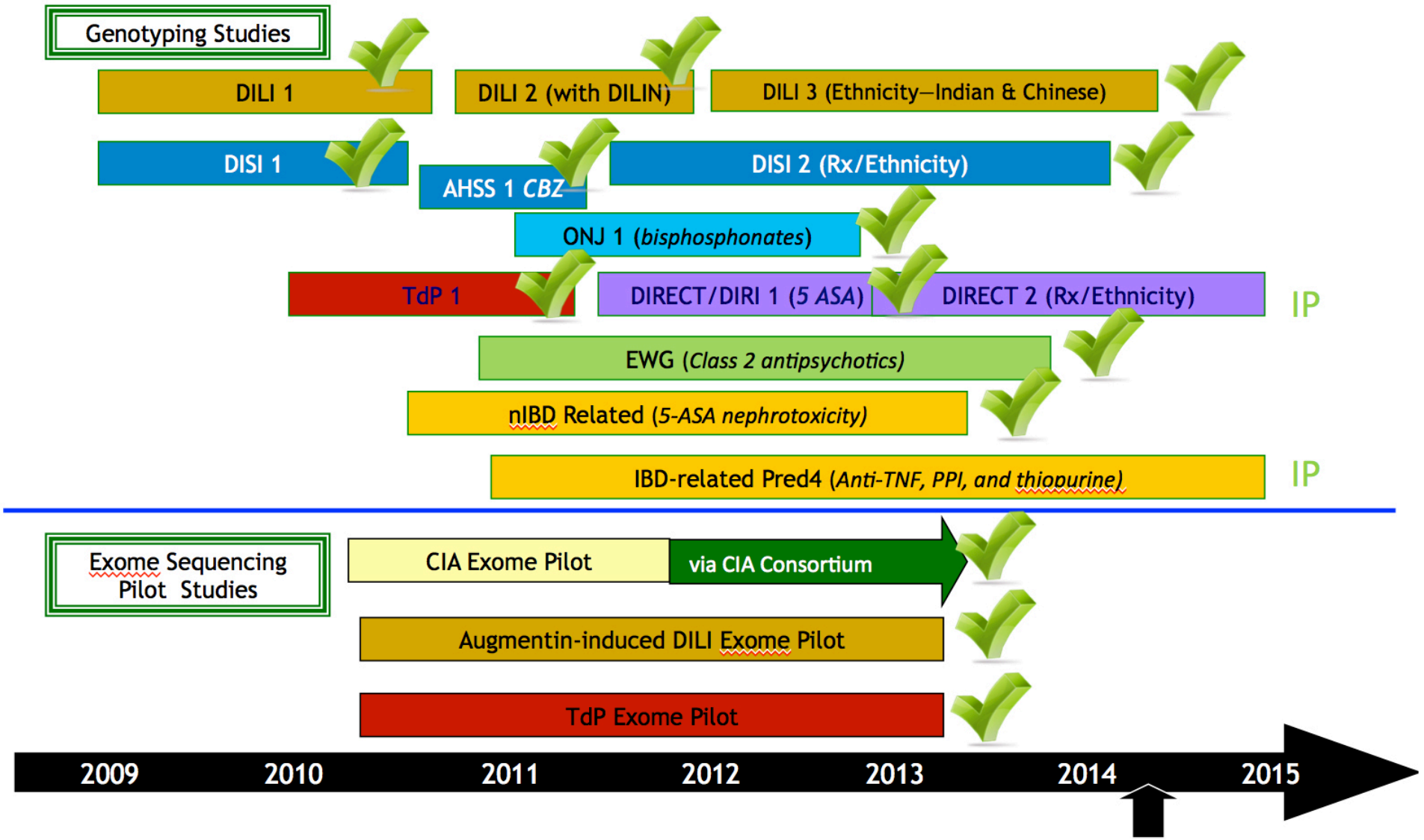
FURTHER INFORMATION
iSAEC Data Access Site: https://dataportal.saeconsortium.org/
iSAEC Public Data Access Policy: http://www.saeconsortium.org/?q=node/27

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (box)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Box S1 | Summary of case enrolment by iSAEC and status of phase 2 pipeline

	Phase 1			Phase 2			Grand totals	%
	iSAEC	Collaborators	Total	iSAEC	Collaborators	Total		
DILI (iDILIC)	505	401	906	1,331	424	1,755	**2,661**	36%
DISI (ITCH)	90	0	90	1,332	0	1,332	**1,422**	19%
DiTdP	74	206	280	-	-	-	**280**	4%
DIRI (DIRECT) forecasted	0	0	0	800	100	900	**900**	12%
BNONJ	0	0	0	358	0	358	**358**	5%
IBD SAE/ADR-related (PRED6)	0	0	0	1,636	-	1,636	**1,636**	22%
Excessive weight gain/CL2APs	0	0	0	175	0	175	**175**	2%
Total	**669**	**607**	**1,276**	**5632**	**524**	**6,156**	**7,432**	

In format provided by Holden *et al.* (NOVEMBER 2014)



Box S2 | Publications from the International Serious Adverse Events Consortium to date (as of September 2014)

- Daly, A. K. *et al.* HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nature Genet.* **41**, 816-819 (2009).
- Lucena, M. I. *et al.* Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class I and II alleles. *Gastroenterology* **141**, 338-347 (2011).
- Shen, Y. *et al.* Genome-wide association study of serious blistering skin rash caused by drugs. *Pharmacogenomics J.* **12**, 96-104 (2012).
- McCormack, M. *et al.* HLA-A*3101 and carbamazepine-induced hypersensitivity reactions in Europeans. *N. Engl. J. Med.* **364**, 1134-1143 (2011).
- Pirmohamed, M. *et al.* The Phenotype Standardization Project: improving pharmacogenetic studies of serious adverse drug reactions. *Clin. Pharmacol. Ther.* **89**, 784-758 (2011).
- Pirmohamed, M. *et al.* Phenotype standardization for immune-mediated drug-induced skin injury. *Clin. Pharmacol. Ther.* **89**, 896-901 (2011).
- Aithal, G. P. *et al.* Case definition and phenotype standardization in drug-induced liver injury. *Clin. Pharmacol. Ther.* **89**, 806-815 (2011).
- Behr, E. R. *et al.* DiTdP à The International Serious Adverse Events Consortium (iSAEC) phenotype standardization for drug-induced torsades de pointes. *Eur. Heart J.* **34**, 1958-1963 (2013).
- Contreras, J. Information access. Prepublication data release, latency, and genome commons. *Science* **329**, 393-394 (2012).
- Behr, E. R. *et al.* Genome wide analysis of drug-induced torsades de pointes: lack of common variants with large effect sizes. *PLoS One* **8**, e78511 (2013).
- Sullivan, P. F. *et al.* Clozapine-induced agranulocytosis is associated with rare *HLA-B* and *HLA-DQB1* alleles". *Nature Comms.* 4 Sep 2014 (doi: 10.1038/ncomms 5757)
- Heap, G. A. *et al.* HLA DQA1-DRB1 variants confer susceptibility to pancreatitis induced by the thiopurine immunosuppressants. *Nature Genet.* 15 Sept 2014 (doi:10.1038/ng.3093)