Review Article

Theme: New Paradigms in Pharmaceutical Sciences: In Silico Drug Discovery
Guest Editors: Xiang-Qun Xie

Paradigm Shift in Toxicity Testing and Modeling

Hongmao Sun,^{1,2} Menghang Xia,¹ Christopher P. Austin,¹ and Ruili Huang^{1,2}

Received 14 September 2011; accepted 5 April 2012; published online 20 April 2012

Abstract. The limitations of traditional toxicity testing characterized by high-cost animal models with low-throughput readouts, inconsistent responses, ethical issues, and extrapolability to humans call for alternative strategies for chemical risk assessment. A new strategy using *in vitro* human cell-based assays has been designed to identify key toxicity pathways and molecular mechanisms leading to the prediction of an *in vivo* response. The emergence of quantitative high-throughput screening (qHTS) technology has proved to be an efficient way to decompose complex toxicological end points to specific pathways of targeted organs. In addition, qHTS has made a significant impact on computational toxicology in two aspects. First, the ease of mechanism of action identification brought about by *in vitro* assays has enhanced the simplicity and effectiveness of machine learning, and second, the high-throughput nature and high reproducibility of qHTS have greatly improved the data quality and increased the quantity of training datasets available for predictive model construction. In this review, the benefits of qHTS routinely used in the US Tox21 program will be highlighted. Quantitative structure–activity relationships models built on traditional *in vivo* data and new qHTS data will be compared and analyzed. In conjunction with the transition from the pilot phase to the production phase of the Tox21 program, more qHTS data will be made available that will enrich the data pool for predictive toxicology. It is perceivable that new *in silico* toxicity models based on high-quality qHTS data will achieve unprecedented reliability and robustness, thus becoming a valuable tool for risk assessment and drug discovery.

KEY WORDS: computational toxicology; qHTS; risk assessment; Tox21.

INTRODUCTION

The gold standard for chemical hazard evaluation has been *in vivo* toxicology, where the compound is formulated and dosed against rodents or other animals (1). The high cost of *in vivo* testing, in terms of time, resource consumption, and animal usage, has confined its usage to testing limited numbers of compounds in a low-throughput manner. It has been estimated that 7,000 animals are needed for each registered pesticide at a cost around tens of millions of dollars in a matter of years (2). Therefore, the reality is that on one hand, more than a thousand new chemicals and pharmaceuticals are produced each year and thus pose a potential threat to the environment and human health, while on the other hand, *in vivo* tests can only evaluate a fraction of them (3). The situation calls for alternative strategies for chemical risk assessment. To advance the state of toxicity testing, the US Tox21 program has initiated a paradigm shift in toxicity testing of chemical compounds from traditional *in vivo* tests to less expensive and higher-throughput cell-based assays in order to identify key pathways and proteins linked with toxicity end points (4). The US Tox21 program is aimed at accelerating the

development of mechanism-based *in vitro* screens to better understand the mechanisms of toxicity and to reduce the use of low-throughput, high-cost traditional toxicity testing relied on animal models (5,6). The subsequent years observed the successful collaborations that brought together the *in vivo* experimental expertise of the National Toxicology Program (NTP), quantitative high-throughput screening (qHTS) technology of the National Institutes of Health Chemical Genomics Center (NCGC), and the computational toxicology capability of the US Environmental Protection Agency (4). Last year, the US Food and Drug Administration joined the US Tox21 program, potentially providing accessibility to human toxicity data for thousands of drugs entered into clinical trials.

Traditional HTS is capable of testing millions of compounds in a week to a month at a single concentration (typically at 10 μ M) at a cost of high false positive (FP) and false negative (FN) rates (7). Certain levels of FP and FN rates for HTS are acceptable as part of a drug discovery project, as long as a few true hits can be confirmed as the starting point of the project. On the contrary, toxicity testing has less tolerance to FPs and especially FNs, since FP results may block or delay a good product from reaching the market, whereas a FN labeling may have serious consequences in regards to the environment or human health. Therefore, a testing system that is more reliable and reproducible than the traditional HTS is highly desirable for the evaluation of adverse effects of chemicals. The qHTS technique, which emerged a few years ago,

¹Department of Health and Human Services, NIH Chemical Genomics Center, National Institutes of Health, 9800 Medical Center Drive, Bethesda, Maryland 20892-3370, USA.

²To whom correspondence should be addressed. (e-mail: hongmao.sun@nih.gov; huangru@mail.nih.gov)

meets the requirements for toxicity testing (8–10), where each compound is tested at multiple (4,7–14) concentrations across a four-log unit concentration span, producing a concentration–response curve, thus effectively reducing the FP and FN rates. The high-quality data generated from qHTS not only consolidates the ground work on which the risk assessments are based but also lays a solid foundation for computational toxicology. In addition, the qHTS technology improves the data quality without sacrificing the throughput and boosting the capability of screening hundreds of thousands of compounds per week. Given the benefits brought about by qHTS, this technology is not free from technical issues common with HTS (11) or from assay artifacts introduced by, e.g., compound autofluorescence, aggregation, and cytotoxicity (12,13).

The toxicity of a chemical can be related directly to its structure; thus, studies of quantitative structure–activity relationships (QSAR) should shed light on understanding the molecular mechanisms of action (MOA) of toxic substances. Machine learning is one of the most frequently used techniques to train computer systems to differentiate actives from inactives within a set of compounds of known activities (14–19). Predictive toxicology has incomparable advantages in terms of throughput, cost, and expandability to virtual compounds (1,20–23); however, reliability and robustness are two major concerns preventing it from being widely applied to risk assessment. The quality of a QSAR model is largely dependent on three factors: data quality, molecular descriptors, and statistical algorithms employed to build the model (24). Statistical methods are relatively matured and no significant improvement is expected for the near future, yet selecting a suitable method from a pool of statistical methods is not trivial. Support vector machines (SVM) (14,17), recursive partitioning (25), random forest (26), and Bayesian statistics (27) are the most frequently used nonlinear algorithms, while partial least squares projection to latent structures (28) is the most popular linear method. Both free and commercial molecular descriptors are currently in abundance for model construction. Those molecular descriptors include whole molecule properties (29), fragment or functional group-based descriptors (30), atom types (31), and molecular fingerprints (32). Molecular descriptors contribute tremendously to the predictive power and robustness of a QSAR model. Data quality is another major determinant of the quality of a model, since “garbage-in, garbage-out” should be avoided in the field of machine learning. The data aspect of modeling includes not only the quality and quantity of a data set but also molecular diversity and the distribution or balance of the responses (24). The newly available qHTS datasets light new hope for computational toxicology by removing the major huddle of data concerns from the path towards reliable and robust toxicity models.

IN VIVO TESTING LIMITS THE EFFICIENCY OF RISK ASSESSMENT

Various animal models have been developed to evaluate acute, subacute, and chronic effects of tested compounds for such end points as oral, dermal, and ocular toxicity, immunotoxicity, genotoxicity, reproductive and developmental

toxicity, and carcinogenicity (9,33–37). *In vivo* animal models offer direct evidence of chemical toxicity in a living organism, but this strength is offset by the weaknesses of low-throughput and excessive demand of resources (4). For example, chemicals under toxicology or carcinogenesis evaluations by the NTP take 14 days (in a subacute study) to 2 years (in a chronic study) of exposure in rodents (38). With tens of thousands of chemicals currently in use, lack of proper safety assessment (39), and about a thousand new chemicals of commercial and environmental importance emerging each year (3), *in vivo* testing is becoming inadequate in order to provide timely, much needed safety information. Furthermore, increasing pressure from animal protection coalitions forces people to rethink the policy of using large quantities of animals in toxicology research. Moreover, inconsistent testing results associated with the intrinsic differences of individual animals may prevent researchers from drawing any definitive conclusions on a test compound, especially for orally dosed chronic studies (40). Lastly, the *in vivo* testing results of rodents are not readily translatable to humans due to differences in species sensitivity (9). In order to bridge the gap between high-dose toxicity in rodents and low-dose risks in humans, physiologically based pharmacokinetics (PBPK), a mathematic modeling technique, may be needed to minimize the cross-species differences in physiology (41,42). PBPK research has exhibited some encouraging capabilities in profiling the human PK of a compound from mouse PK data; however, PBPK is still a long way away from being able to deliver consistent predictions (43).

ROBOTIC CELL-BASED QHTS PLATFORM DELIVERS HIGH THROUGHPUT AND HIGH QUALITY

The HTS technique, originally developed by the pharmaceutical industry to identify bioactive compounds from huge corporate compound libraries, greatly enhanced the throughput of biological assays (44,45). The trade-off of high throughput of the traditional single concentration HTS is its high FP and FN rates (7). In order to overcome this shortcoming of traditional HTS, a new technology, quantitative HTS, was developed by the NCGC (46) and has matured and been widely applied in subsequent years (8,9,13,47–55). The cytotoxicity profiling of the NTP library of 1,408 environmental chemicals against 13 human and rodent cells and cell lines clearly exhibited the robustness and reproducibility of the qHTS technique, where three independent runs on the HepG2 cells produced a high correlation of $r^2=0.87$ (9). In addition to the inherent benefits of greatly improved data quality, qHTS can either use various human cell lines to make the assay results more relevant to human toxicology or use rodent cell lines to improve the understanding of relationships between *in vitro* data and *in vivo* rodent toxicity accumulated over the last two decades. Another advantage of using cell-based assays is that the permeability of testing compounds is already taken into account. By combining the cell-based toxicity data with the PK profile of the tested compounds (56), their oral toxicity might be reliably predicted.

The Tox21 program has acquired a new robotic system (Fig. 1), which is located at NCGC, and is expected to greatly

Fig. 1. The Tox21 robotic system. There are a series of peripherals and workstations arranged around a central anthropomorphic robotic arm. The main components include the Kalypsys Director software, a pin tool device for nanoliter compound transfer, plate storage and environmentally controlled assay incubation units, nanoliter reagent dispensers (BioRPTR, Multidrop, acoustic/noncontact (ATS-100)), a centrifuge (V-spin), and plate readers (ViewLux and EnVision)

accelerate the transition of the Tox21 program from its pilot phase to its production phase. During the production phase, over 10,000 chemicals, of which approximately two thirds are environmental chemicals and the rest pharmaceuticals, will be screened against a wide spectrum of assays in order to cover as many known relevant pathways as possible for toxicity testing.

THE HISTORY OF COMPUTATIONAL TOXICOLOGY

It has long been recognized that the biological activity, including toxicity of a chemical, resides somewhere in the structure of the chemical (57), yet it continues to remain a challenge to fully understand the mechanism of toxicity at the cellular and molecular level. In the early 1990s, the National Institute of Environmental Health Sciences initiated a Predictive Toxicology Evaluation (PTE) project, aiming at a rigorous and objective evaluation on the performance of predictive toxicology (58). In the first round of the PTE experiment (PTE-1), the PT community was challenged to predict the carcinogenicity of 44 chemicals, and in the following PTE-2 practice, a set of 30 different compounds were supplied by the NTP for prediction (59). The training set of 313 NTP chemicals were classified for each sex-species experiment into four overall levels of evidence, namely clear evidence, some evidence, equivocal evidence, and no evidence, according to the long-term chemical carcinogenesis

experiments (60). The PTE results were intensely reviewed (61–66), and some consensus was reached, including QSAR-based models were outperformed by models using biological attributes (58). Consequently, the best QSAR model using inductive logic programming could only achieve a predictive accuracy of 63 % in a 5-fold cross validation (67). Why did QSAR approaches fail in predicting chemical carcinogenesis?

PROBLEM COMPLEXITY AND APPLICABILITY DOMAIN

One of the most frequently asked questions in QSAR research is how large a training set is needed to ensure a reasonable predictive model. The answer to this question depends on many factors, among which the most important determinant is the complexity of the concept to learn.

QSAR is a special case of machine learning. Machine learning is to train computer systems with existing knowledge, such as measured toxicity of compounds in a training set, and the learners then make predictions on the basis of the experience built over the learning materials. The concept to learn has different levels of complexity, as demonstrated in Fig. 2; thus, the size of a training set required to learn these concepts also varies. Two data points can define a straight line in space (Fig. 2a); a few more data points may be needed to reinforce the linear assumption without considering errors. Nonlinear responses add more difficulties to the model, but a

Fig. 2. a–c Graphics illustrating the increasing complexity level of concepts

smart learner can figure out the relationships that are hidden in a dataset governing the concept through transforming a nonlinear problem to a linear one in feature space, as long as the problem is analytically solvable (Fig. 2b). A nonlinear problem with a high level of random noise (Fig. 2c) requires a larger training data set in order to be simulated. Biological data are usually nonlinear, multi-dimensional, and can be highly variable from experiment to experiment (68); therefore, it is less likely to achieve a reasonably reliable and robust QSAR model with less than a thousand balanced training compounds, although the optimum size of a training set also depends on the types of descriptors and statistical methods (69).

Mechanisms of chemical carcinogenicity have been thoroughly investigated, and a considerable amount of research has been published over the past few decades (70). Chemical carcinogens can be divided into genotoxic (or DNA-reactive) carcinogens and nongenotoxic carcinogens. Many classes of chemical carcinogens, such as alkylating agents, acylating agents, and arylaminating agents, have been found to react with different sites of DNA bases, but not all DNA-reactive chemicals are carcinogens, since cells possess a variety of DNA repair processes (71–73). Mechanisms of nongenotoxic carcinogens are also complicated, which include peroxisome proliferator effects (74), cytotoxicity, immunosuppression, induction of hormonal imbalance, and oxidative stress (70). In conclusion, chemical carcinogenicity is extremely complicated for machine learning. As a result, a training set of 313 compounds in the PTE experiment is far from sufficient for the construction of any predictive model. Any attempt to learn a complex concept with limited training sets can lead to failure, which can partially explain the failure of the QSAR modeling efforts in PTE-1 and PTE-2.

One way to overcome the local effects associated with limited training data is to build a local QSAR model and to confine its usage to the coverage of the training data that the model is based on. As illustrated in Fig. 3, a simpler concept may be embedded in an overall complicated concept. In such a case, a local model can be valid as long as it is applied within its applicability domain (AD), defined by the structural space spanned by the training data (75). Once the AD of a QSAR model is defined, the accuracy or confidence of its predictions can be estimated by the distance to model of individual test compounds (76,77). To prevent predictive errors introduced by extrapolations, one should avoid

applying a local model outside of its AD. Multiple local models might be built to cover different regions of a complex concept, but misclassifying a test sample to the wrong local model may ruin the predictive performance. Another strategy to deal with complex concepts is to simplify them through decomposition.

DECOMPOSE COMPLEX MOA TO PATHWAYS AND ENZYMES

Assuming a compound exhibits adverse effects by interacting with molecular targets A, B, or C, subsequently the QSAR model has to cover the chemical space representing the ligand binding sites of three different targets. This kind of QSAR model generally requires a larger training dataset than individual QSAR models for each of the three targets (78). The dilemma here is that a complex MOA requires a large training set to decipher the underlying complexity, whereas the data available for complex MOAs, such as *in vivo* toxicity, for training purposes are limited (62).

Most of the toxicological end points are entangled with multiple mechanisms, pathways, and enzymes, as discussed in carcinogenicity. The complex toxicological end points can often be broken down into a few specific pathways in specific

Fig. 3. A picture elucidating the concept of applicability domain (AD) and the increased deviation outside of AD

target organs (57). The more detailed and specific information on the molecular targets behind the toxicity is revealed, the more likely the MOA of the chemical toxicity can be understood. The US Tox21 program aims to identify such *in vitro* assays that can detect perturbations in cellular pathways that are expected to contribute to or result in adverse effects (4,79). To decipher chemical toxicity at the cellular and molecular level, the Tox21 program has established in its pilot phase a number of pathway- or target-based assays with initial focuses on stress response and nuclear receptor signaling pathways. The selection of stress response pathway assays is based on the premise that compounds inducing one or more stress response pathways are more likely to exhibit *in vivo* toxicity than those that do not (80). Examples of such stress response pathway assays include assays that can detect compounds capable of inducing oxidative stress by activating the antioxidant response element via the Nrf2 signaling pathway (Shukla *et al.*, MS in preparation) (81), DNA damage response by inducing differential cytotoxicity in several different DNA repair deficient DT40 chicken isogenic cell clones compared to the wild-type cells (82), mitochondrial damage by causing changes in mitochondrial membrane potential (Sakamuru *et al.*, MS in preparation), inflammatory response by activating the NF κ B (Toll-like receptor) signaling pathway (83) or inducing TNF α production (47), hypoxic response by activating hypoxia inducible factor-1 alpha activity (84), cardiotoxicity by inhibiting the human ether- α -go-go-related gene (hERG) potassium channel (85), and apoptosis by activating the caspase-3/7 pathway (54). Endocrine disruption-related toxicity is a major concern for environmental chemicals. The Tox21 pilot phase collection of approximately 3,000 compounds has been screened against ten human nuclear receptors including ER α , AR, GR, TR β , PPAR δ , PPAR γ , FXR, LXR, RXR, and VDR (13).

IMPROVED PREDICTIVE PERFORMANCE ACHIEVED BY USING QHTS DATA

The qHTS technology has not only helped to increase the quantity but also improve the quality of biological data, which has deterred the development of reliable QSAR models (24,86). The qHTS technology, combining the strength of higher-throughput and lower experimental errors, has provided the QSAR community with several good quality datasets, including the inhibition data of five isozymes of human cytochrome P450 (CYP450) (87), hERG (85), cytotoxicity, etc. These valuable datasets are publicly available through PubChem (88,89). QSAR models built on the basis of the qHTS data assumed higher reliability and better coverage of the chemical space (18,90–92). The SVM classification models trained using the CYP450 data achieved high predictive accuracies of between 0.85 and 0.93 over the equal-sized test sets for all five CYP450 isozymes, as measured by the area under the receiver operating characteristic curves (AUC-ROC) (92). The CYP450 models constructed from over 6,000 drug-like small molecules exhibited high predictive accuracies of AUC-ROC (between 0.78 and 0.85) when used to predict a collection of non-drug-like environmental toxins, such as pollutants, food additives, and pesticides (Sun *et al.*, submitted for publication). The

unexpected high predictive accuracies over the dissimilar compounds can be attributed largely to the high quantity of qHTS training data measured under the same conditions and atom typing descriptors, which greatly expanded the coverage of chemical space and the AD of the resulting models.

RECONNECT *IN VITRO* DATA AND *IN VIVO* TOXICITY

In vitro data generated via qHTS campaigns are tied to specific cell types and/or pathways; thus, these activities may or may not be directly related to any toxicological end points, which is more meaningful to public health. In order to reconnect *in vitro* data with *in vivo* toxicities, many efforts have been made over the past few years. In order to improve the efficiency of detecting toxicity mechanisms, we (54) measured the cell viability and caspase activation of 13 different cell types against the NTP library of 1,408 compounds. The assay results were later analyzed using an unsupervised *k*-means clustering algorithm to maximize Dunn's index, a measure of cluster quality in terms of internal consistency (93). The results suggested that a set of six assays could achieve 94 % of the clustering resolution achieved by all 26 assays (54). By removing redundant assays, the same compound MOA can be reached at a significantly reduced cost. Due to our limited understanding of toxicity mechanisms, it is difficult to determine *a priori* which assays are relevant for a given toxicity end point. In another attempt, DiMaggio *et al.* biclustered the *in vitro* assays by iterative optimal reordering of the 309 ToxCast compounds and the 524 assays in a heat map to determine whether there existed correlative changes in the chemicals over particular subsets of assays (94). The biclustering practice on *in vitro* data produced reasonable clusters on assay targets and types. The sparse *in vivo* data set was processed similarly, and the toxicity end points were observed to cluster according to animal species and similar physiological attributes (94). Then the *in vitro* and *in vivo* data were cross-linked via regularized logistic regression, resulting in the conclusion that the CYP assays exhibited a specificity for the liver end points while estrogen/androgen nuclear receptors were indicators of the reproductive end points (94). These results suggest a promising opportunity to reconnect *in vitro* and *in vivo* assays through a set of compounds without any *a priori* knowledge of assay targets and toxicity end points; however, quantitative relationships between *in vitro* and *in vivo* results cannot be derived using this approach. *In vitro* data can also be used in combination with chemical descriptors to enhance the predictability of a QSAR model. Tropsha's group observed that the predictive accuracy of rodent carcinogenicity using *k*-nearest neighbor (*k*NN) QSAR models significantly improved to 72.7 % when combining chemical descriptors with human cell viability assay data from the 62.3 % obtained using chemical descriptors alone (95). The same group included the cell viability qHTS concentration–response data in 13 cell lines as biological descriptors when constructing QSAR models, and a similar trend of improving model accuracy was obtained in their *k*NN and random forest QSAR models for the prediction of rodent acute toxicity (96). As these approaches have achieved limited success in making use of *in vitro* assay data to enhance our understanding of *in vivo* toxicity, significantly more work is needed to meet the ultimate goal, i.e., to design a set of *in vitro*

assays to produce results that can be used as good predictors of *in vivo* toxicities. As a result of the Tox21 program entering its production phase, more cell-based qHTS data will become available to the public domain, and the gap between the *in vitro* assays and *in vivo* activities will be mitigated gradually.

SUMMARY

Rapid advances in molecular and cellular biology in the last two decades declare the advent of a new era in chemical toxicity testing. The newly emerged qHTS technology enables the capacity of testing tens of thousands of compounds per week with unprecedented quality and reproducibility. The US Tox21 program routinely screens chemicals of environmental and human health importance using cell-based qHTS assays, aiming to identify key pathways or molecular targets relevant to toxicity. The qHTS assays have shed light into understanding the events leading to toxic responses at the level of cellular and molecular interactions and reactions. The qHTS assay results of significantly improved quality and quantity will pave the road toward robust and reliable QSAR models. The QSAR models based on the qHTS data have demonstrated higher predictive accuracies and broader coverage of chemical space. Computational toxicology plays an important role not only in predicting the toxicity of a compound but also in prioritizing compounds for further testing, systematically evaluating in-depth toxicology, suggesting safer chemicals to use in the environment, in food, or in the pharmaceutical industry, and also in elucidating the mechanisms by which chemicals exert their toxic action on living organisms. Primary successes have been achieved on modeling *in vivo* toxicity using *in vitro* assay results and on cross-linking *in vivo* and *in vitro* assay results. With the Tox21 program entering its production phase, we are closer to our goal of reducing and ultimately replacing animal based *in vivo* testing with cost efficient *in vitro* assays for chemical risk assessment.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Programs (Interagency agreement #Y2-ES-7020-01) of the National Toxicology Program, National Institute of Environmental Health Sciences (NIEHS), and the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH). The statements, opinions, or conclusions contained therein do not necessarily represent the statements, opinions, or conclusions of NIEHS, or NCATS, NIH, or the US government. We thank in particular Anna Rossoshek for helpful comments and suggestions during the preparation of this manuscript.

REFERENCES

- Merlot C. Computational toxicology—a tool for early safety evaluation. *Drug Discov Today*. 2010;15(1–2):16–22.
- Erickson BE. Modernizing toxicity tests. *Chem Eng News*. 2011;89:25–6.
- Tetko IV, Sushko I, Pandey AK, Zhu H, Tropsha A, Papa E, *et al*. Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection. *J Chem Inf Model*. 2008;48(9):1733–46.
- Shukla SJ, Huang R, Austin CP, Xia M. The future of toxicity testing: a focus on *in vitro* methods using a quantitative high-throughput screening platform. *Drug Discov Today*. 2010;15(23–24):997–1007.
- Schmidt CW. TOX 21: new dimensions of toxicity testing. *Environ Health Perspect*. 2009;117(8):A348–53.
- Kavlock RJ, Austin CP, Tice RR. Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal*. 2009;29(4):485–7. discussion 92–7.
- Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R. Statistical practice in high-throughput screening data analysis. *Nat Biotechnol*. 2006;24(2):167–75.
- Johnson RL, Huang R, Jadhav A, Southall N, Wichterman J, MacArthur R, *et al*. A quantitative high-throughput screen for modulators of IL-6 signaling: a model for interrogating biological networks using chemical libraries. *Mol Biosyst*. 2009;5(9):1039–50.
- Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho MH, *et al*. Compound cytotoxicity profiling using quantitative high-throughput screening. *Environ Health Perspect*. 2008;116(3):284–91.
- Houck KA, Dix DJ, Judson RS, Kavlock RJ, Yang J, Berg EL. Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. *J Biomol Screen*. 2009;14(9):1054–66.
- Yasgar A, Shinn P, Jadhav A, Auld D, Michael S, Zheng W, *et al*. Compound management for quantitative high-throughput screening. *JALA Charlottesville Va*. 2008;13(2):79–89.
- Jadhav A, Ferreira RS, Klumpp C, Mott BT, Austin CP, Inglese J, *et al*. Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease. *J Med Chem*. 2010;53(1):37–51.
- Huang R, Xia M, Cho MH, Sakamuru S, Shinn P, Houck KA, *et al*. Chemical genomics profiling of environmental chemical modulation of human nuclear receptors. *Environ Health Perspect*. 2011;119(8):1142–8.
- Cristianini N, Shawe-Taylor J. An introduction to support vector machines. Cambridge: Cambridge University Press; 2005.
- Fox T, Kriegl JM. Machine learning techniques for *in silico* modeling of drug metabolism. *Curr Top Med Chem*. 2006;6(15):1579–91.
- Judson R, Elloumi F, Setzer RW, Li Z, Shah I. A comparison of machine learning algorithms for chemical toxicity classification using a simulated multi-scale data model. *BMC Bioinforma*. 2008;9:241.
- Noble WS. What is a support vector machine? *Nat Biotechnol*. 2006;24(12):1565–7.
- Novotarskyi S, Sushko I, Korner R, Pandey AK, Tetko IV. A comparison of different QSAR approaches to modeling CYP450 1A2 inhibition. *J Chem Inf Model*. 2011;51(6):1271–80.
- Schroeter T, Schwaighofer A, Mika S, Laak AT, Suelzle D, Ganzer U, *et al*. Machine learning models for lipophilicity and their domain of applicability. *Mol Pharm*. 2007;4(4):524–38.
- Kavlock RJ, Ankley G, Blacato J, Breen M, Conolly R, Dix D, *et al*. Computational toxicology—a state of the science mini review. *Toxicol Sci*. 2008;103(1):14–27.
- Nigsch F, Macaluso NJ, Mitchell JB, Zmuidinavicius D. Computational toxicology: an overview of the sources of data and of modelling methods. *Expert Opin Drug Metab Toxicol*. 2009;5(1):1–14.
- Muster W, Breidenbach A, Fischer H, Kirchner S, Muller L, Pahler A. Computational toxicology in drug development. *Drug Discov Today*. 2008;13(7–8):303–10.
- Benfenati E, Gini G. Computational predictive programs (expert systems) in toxicology. *Toxicology*. 1997;119(3):213–25.
- Cronin MTD, Schultz TW. Pitfalls in QSAR. *J Mol Struct Theochem*. 2003;622(1–2):39–51.
- Cook EF, Goldman L. Empiric comparison of multivariate analytic techniques: advantages and disadvantages of recursive partitioning analysis. *J Chron Dis*. 1984;37(9–10):721–31.
- Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
- Berger JO. Statistical decision theory and Bayesian analysis. New York: Springer; 1985.
- Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab*. 2001;58(2):109–30.
- Helguera AM, Combes RD, Gonzalez MP, Cordeiro MN. Applications of 2D descriptors in drug design: a DRAGON tale. *Curr Top Med Chem*. 2008;8(18):1628–55.

30. Shen J, Cheng F, Xu Y, Li W, Tang Y. Estimation of ADME properties with substructure pattern recognition. *J Chem Inf Model*. 2010;50(6):1034–41.
31. Crippen GM, Wildman SA. Prediction of physicochemical parameters by atomic contributions. *J Chem Inf Comput Sci*. 1999;39(5):868–73.
32. Butina D. Unsupervised data base clustering based on Daylight's fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets. *J Chem Inf Comput Sci*. 1999;39(4):747–50.
33. Buse E, Habermann G, Osterburg I, Korte R, Weinbauer GF. Reproductive/developmental toxicity and immunotoxicity assessment in the nonhuman primate model. *Toxicology*. 2003;185(3):221–7.
34. Farine JC. Animal models in autoimmune disease in immunotoxicity assessment. *Toxicology*. 1997;119(1):29–35.
35. Trobridge GD, Kiem HP. Large animal models of hematopoietic stem cell gene therapy. *Gene Ther*. 2010;17(8):939–48.
36. Penha FM, Rodrigues EB, Maia M, Dib E, Fiod Costa E, Furlani BA, *et al*. Retinal and ocular toxicity in ocular application of drugs and chemicals—part I: animal models and toxicity assays. *Ophthalmic Res*. 2010;44(2):82–104.
37. Perlman I. Testing retinal toxicity of drugs in animal models using electrophysiological and morphological techniques. *Doc Ophthalmol*. 2009;118(1):3–28.
38. Chhabra RS, Bucher JR, Wolfe M, Portier C. Toxicity characterization of environmental chemicals by the US National Toxicology Program: an overview. *Int J Hyg Environ Health*. 2003;206(4–5):437–45.
39. Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, *et al*. The toxicity data landscape for environmental chemicals. *Environ Health Perspect*. 2009;117(5):685–95.
40. Gottmann E, Kramer S, Pfahringer B, Helma C. Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. *Environ Health Perspect*. 2001;109(5):509–14.
41. Marino DJ. Physiologically based pharmacokinetic modeling using Microsoft Excel and Visual Basic for applications. *Toxicol Mech Methods*. 2005;15(2):137–54.
42. Yang RS, Thomas RS, Gustafson DL, Campain J, Benjamin SA, Verhaar HJ, *et al*. Approaches to developing alternative and predictive toxicology based on PBPK/PD and QSAR modeling. *Environ Health Perspect*. 1998;106 Suppl 6:1385–93.
43. Germani M, Crivori P, Rocchetti M, Burton PS, Wilson AG, Smith ME, *et al*. Evaluation of a basic physiologically based pharmacokinetic model for simulating the first-time-in-animal study. *Eur J Pharm Sci*. 2007;31(3–4):190–201.
44. Macarron R. Critical review of the role of HTS in drug discovery. *Drug Discov Today*. 2006;11(7–8):277–9.
45. Gribbon P, Sewing A. High-throughput drug discovery: what can we expect from HTS? *Drug Discov Today*. 2005;10(1):17–22.
46. Inglesse J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, *et al*. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc Natl Acad Sci USA*. 2006;103(31):11473–8.
47. Leister KP, Huang R, Goodwin BL, Chen A, Austin CP, Xia M. Two high throughput screen assays for measurement of TNF- α in THP-1 cells. *Curr Chem Genomics*. 2011;5:21–9.
48. Shukla SJ, Sakamuru S, Huang R, Moeller TA, Shinn P, Vanleer D, *et al*. Identification of clinically used drugs that activate pregnane X receptors. *Drug Metab Dispos*. 2011;39(1):151–9.
49. Xia M, Guo V, Huang R, Inglesse J, Nirenberg M, Austin CP. A cell-based beta-lactamase reporter gene assay for the CREB signaling pathway. *Curr Chem Genomics*. 2009;3(1):7–12.
50. Thomas CJ, Auld DS, Huang R, Huang W, Jadhav A, Johnson RL, *et al*. The pilot phase of the NIH chemical genomics center. *Curr Top Med Chem*. 2009;9(13):1181–93.
51. Titus SA, Beacham D, Shahane SA, Southall N, Xia M, Huang R, *et al*. A new homogeneous high-throughput screening assay for profiling compound activity on the human ether-a-go-go-related gene channel. *Anal Biochem*. 2009;394(1):30–8.
52. Xia M, Huang R, Guo V, Southall N, Cho MH, Inglesse J, *et al*. Identification of compounds that potentiate CREB signaling as possible enhancers of long-term memory. *Proc Natl Acad Sci USA*. 2009;106(7):2412–7.
53. Cho MH, Niles A, Huang R, Inglesse J, Austin CP, Riss T, *et al*. A bioluminescent cytotoxicity assay for assessment of membrane integrity using a proteolytic biomarker. *Toxicol In Vitro*. 2008;22(4):1099–106.
54. Huang R, Southall N, Cho MH, Xia M, Inglesse J, Austin CP. Characterization of diversity in toxicity mechanism using *in vitro* cytotoxicity assays in quantitative high throughput screening. *Chem Res Toxicol*. 2008;21(3):659–67.
55. Inglesse J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, *et al*. High-throughput screening assays for the identification of chemical probes. *Nat Chem Biol*. 2007;3(8):466–79.
56. Di L, Kerns EH. Application of pharmaceutical profiling assays for optimization of drug-like properties. *Curr Opin Drug Discov Dev*. 2005;8(4):495–504.
57. McKinney JD. The molecular basis of chemical toxicity. *Environ Health Perspect*. 1985;61:5–10.
58. Bristol DW, Wachsman JT, Greenwell A. The NIEHS predictive-toxicology evaluation project. *Environ Health Perspect*. 1996;104 Suppl 5:1001–10.
59. Benigni R, Andreoli C, Zito R. Prediction of rodent carcinogenicity of further 30 chemicals bioassayed by the U.S. National Toxicology Program. *Environ Health Perspect*. 1996;104 Suppl 5:1041–4.
60. Huff J, Haseman J. Long-term chemical carcinogenesis experiments for identifying potential human cancer hazards: collective database of the National Cancer Institute and National Toxicology Program (1976–1991). *Environ Health Perspect*. 1991;96:23–31.
61. Ashby J. The NIEHS predictive-toxicology evaluation project: the need to distinguish informed uncertainty from ignorant equivocation. *Environ Health Perspect*. 1997;105(7):688.
62. Ashby J, Tennant RW. Prediction of rodent carcinogenicity for 44 chemicals: results. *Mutagenesis*. 1994;9(1):7–15.
63. Tennant RW, Spalding J, Stasiewicz S, Ashby J. Prediction of the outcome of rodent carcinogenicity bioassays currently being conducted on 44 chemicals by the National Toxicology Program. *Mutagenesis*. 1990;5(1):3–14.
64. Ashby J, Tennant RW. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutat Res*. 1991;257(3):229–306.
65. Lee Y, Buchanan BG, Rosenkranz HS. Carcinogenicity predictions for a group of 30 chemicals undergoing rodent cancer bioassays based on rules derived from subchronic organ toxicities. *Environ Health Perspect*. 1996;104 Suppl 5:1059–63.
66. Parry JM. Detecting and predicting the activity of rodent carcinogens. *Mutagenesis*. 1994;9(1):3–5.
67. King RD, Srinivasan A. Prediction of rodent carcinogenicity bioassays from molecular structure using inductive logic programming. *Environ Health Perspect*. 1996;104 Suppl 5:1031–40.
68. Sams-Dodd F. Target-based drug discovery: is something wrong? *Drug Discov Today*. 2005;10(2):139–47.
69. Roy K, Roy PP, Leonard JT. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometr Intell Lab*. 2008;90(1):31–42.
70. Benigni R, Bossa C. Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. *Chem Rev*. 2011;111(4):2507–36.
71. Schwarz T, Schwarz A. DNA repair and cytokine responses. *J Invest Dermatol Symp Proc*. 2009;14(1):63–6.
72. Iyer RR, Pluciennik A, Burdett V, Modrich PL. DNA mismatch repair: functions and mechanisms. *Chem Rev*. 2006;106(2):302–23.
73. Saxowsky TT, Doetsch PW. RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? *Chem Rev*. 2006;106(2):474–88.
74. Gonzalez FJ, Peters JM, Cattley RC. Mechanism of action of the nongenotoxic peroxisome proliferators: role of the peroxisome proliferator-activator receptor α . *J Natl Cancer Inst*. 1998;90(22):1702–9.
75. Jaworska J, Nikolova-Jeliazkova N, Aldenberg T. QSAR applicability domain estimation by projection of the training set descriptor space: a review. *Altern Lab Anim*. 2005;33(5):445–59.
76. Sushko I, Novotarskyi S, Korner R, Pandey AK, Cherkasov A, Li J, *et al*. Applicability domains for classification problems:

- benchmarking of distance to models for Ames mutagenicity set. *J Chem Inf Model*. 2010;50(12):2094–111.
77. Weaver S, Gleeson MP. The importance of the domain of applicability in QSAR modeling. *J Mol Graph Model*. 2008;26(8):1315–26.
  78. Wassermann AM, Peltason L, Bajorath J. Computational analysis of multi-target structure–activity relationships to derive preference orders for chemical modifications toward target selectivity. *ChemMedChem*. 2010;5(6):847–58.
  79. Andersen ME, Krewski D. Toxicity testing in the 21st century: bringing the vision to life. *Toxicol Sci*. 2009;107(2):324–30.
  80. Simmons SO, Fan CY, Ramabhadran R. Cellular stress response pathway system as a sentinel ensemble in toxicological screening. *Toxicol Sci*. 2009;111(2):202–25.
  81. Simmons SO, Fan CY, Yeoman K, Wakefield J, Ramabhadran R. NRF2 oxidative stress induced by heavy metals is cell type dependent. *Curr Chem Genomics*. 2011;5:1–12.
  82. Yamamoto KN, Hirota K, Kono K, Takeda S, Sakamuru S, Xia M, *et al*. Characterization of environmental chemicals with potential for DNA damage using isogenic DNA repair-deficient chicken DT40 cell lines. *Environ Mol Mutagen*. 2011;52(7):547–61.
  83. Miller SC, Huang R, Sakamuru S, Shukla SJ, Attene-Ramos MS, Shinn P, *et al*. Identification of known drugs that act as inhibitors of NF-kappaB signaling and their mechanism of action. *Biochem Pharmacol*. 2010;79(9):1272–80.
  84. Xia M, Huang R, Sun Y, Semenza GL, Aldred SF, Witt KL, *et al*. Identification of chemical compounds that induce HIF-1alpha activity. *Toxicol Sci*. 2009;112(1):153–63.
  85. Xia M, Shahane SA, Huang R, Titus SA, Shum E, Zhao Y, *et al*. Identification of quaternary ammonium compounds as potent inhibitors of hERG potassium channels. *Toxicol Appl Pharmacol*. 2011;252(3):250–8.
  86. Stouch TR, Kenyon JR, Johnson SR, Chen XQ, Doweiko A, Li Y. In silico ADME/Tox: why models fail. *J Comput Aided Mol Des*. 2003;17(2–4):83–92.
  87. Veith H, Southall N, Huang R, James T, Fayne D, Artemenko N, *et al*. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat Biotechnol*. 2009;27(11):1050–5.
  88. Li Q, Cheng T, Wang Y, Bryant SH. PubChem as a public resource for drug discovery. *Drug Discov Today*. 2010;15(23–24):1052–7.
  89. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res*. 2009;37:W623–33. Web Server issue.
  90. Cheng F, Yu Y, Shen J, Yang L, Li W, Liu G, *et al*. Classification of cytochrome P450 inhibitors and noninhibitors using combined classifiers. *J Chem Inf Model*. 2011;51:996–1011.
  91. Shen MY, Su BH, Esposito EX, Hopfinger AJ, Tseng YJ. A comprehensive support vector machine binary hERG classification model based on extensive but biased end point hERG data sets. *Chem Res Toxicol*. 2011;24(6):934–49.
  92. Sun H, Veith H, Xia M, Austin CP, Huang R. Predictive models for CYP450 isozymes based on qHTS data. *J Chem Inf Model*. 2011;51:2474–81.
  93. Bezdek JC, Pal NR. Some new indexes of cluster validity. *IEEE Trans Syst Man Cybern B Cybern*. 1998;28(3):301–15.
  94. DiMaggio Jr PA, Subramani A, Judson RS, Floudas CA. A novel framework for predicting *in vivo* toxicities from *in vitro* data using optimal methods for dense and sparse matrix reordering and logistic regression. *Toxicol Sci*. 2010;118(1):251–65.
  95. Zhu H, Rusyn I, Richard A, Tropsha A. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure–activity relationship models of animal carcinogenicity. *Environ Health Perspect*. 2008;116(4):506–13.
  96. Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, *et al*. Use of *in vitro* HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of *in vivo* toxicity. *Environ Health Perspect*. 2011;119(3):364–70.