

KNOWLEDGE GRAPH AIDS COMPREHENSIVE EXPLANATION OF DRUG TOXICITY

Yun Hao¹, Joseph D. Romano^{2,3}, and Jason H. Moore^{*4}

¹Genomics and Computational Biology (GCB) Graduate Program, University of Pennsylvania, Philadelphia, PA

²Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA

³Center of Excellence in Environmental Toxicology, University of Pennsylvania, Philadelphia, PA

⁴Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA

*To whom correspondence should be addressed: jason.moore@csmc.edu

Keywords: Drug toxicity; Deep learning; Model interpretation; Molecular toxicology

ABSTRACT

In computational toxicology, prediction of complex endpoints has always been challenging, as they often involve multiple distinct mechanisms. State-of-the-art models are either limited by low accuracy, or lack of interpretability due to their black-box nature. Here we introduce AIDTox, an interpretable deep learning model which incorporates curated knowledge of chemical-gene connections, gene-pathway annotations, and pathway hierarchy. AIDTox accurately predicts cytotoxicity outcomes in HepG2 and HEK293 cells. It also provides comprehensive explanations of cytotoxicity covering multiple aspects of drug activity including target interaction, metabolism, and elimination. In summary, AIDTox provides a computational framework for unveiling cellular mechanisms for complex toxicity endpoints.

INTRODUCTION

In silico toxicity assessment is becoming a critical step in drug development owing to sustained advances of machine learning techniques¹. A prevalent approach for toxicity prediction is quantitative structure-activity relationship (QSAR) modeling, which links a toxicity outcome to structural properties of compounds². Coupled with millions of data points generated by large-scale toxicity testing^{3,4}, various QSAR models have been proposed to predict *in vitro* and *in vivo* endpoints⁵⁻¹³. While some have achieved decent predictive performance, almost none can overcome the trade-off between accuracy and interpretability¹⁴. Under state-of-the-art models, it remains challenging to explain the toxicity outcomes of a compound with cellular activities involving target proteins, specific pathways, and biological processes. This limitation has raised substantial concerns among experimental toxicologists, calling for prediction models that can provide insight into cellular mechanisms of toxicity.

Recent developments in visible neural networks (VNN) provide a solution to the issue¹⁵⁻¹⁷. In contrast to black-box neural networks, connections within VNNs are guided by curated knowledge from pathway ontologies. Specifically, the fully connected structure is replaced by an interpretable ontological hierarchy that connects input gene features to output response via hidden pathway modules. In a previous study, we developed a VNN model—named DTox—for predicting compound response to toxicity assays¹⁸. Importantly, DTox advances an innovative interpretation framework that identifies network paths connecting genes and pathways for explaining the toxicity outcome of compounds. We demonstrated that DTox can achieve the same level of predictive performance as state-of-the-art models with a significant improvement in interpretability, as the identified paths can be linked to well-established mechanisms of toxicity. However, one major limitation of DTox is that the input feature profile is derived from structure-based binding prediction models. While these models ensure the wide applicability of DTox, they are vulnerable to prediction errors and data scarcity, causing exclusion of certain genes from the feature space. To address this limitation, we used curated knowledge from the toxicology-focused graph knowledge base ComptoxAI¹⁹ to refine the input feature profile of DTox. ComptoxAI provides extensive profiling of chemical-gene connections across multiple gene categories, which can be used for feature profile construction. Hence, the resulting model—named AIDTox (ComptoxAI + DTox)—contains many novel gene features with active roles in cellular mechanisms of toxicity, including tubulin proteins that regulate apoptosis signaling, cytochrome P450 enzymes that are mainly responsible for drug metabolism, and transporters that participate in drug elimination. We believe AIDTox will facilitate the generation of new hypotheses for mechanistic investigation. Our code can be accessed openly at <https://github.com/yhao-compbio/AIDTox>.

METHODS

Processing cell viability screening datasets for model training

The Tox21 datasets³ contain screening results indicating the response of *in vitro* toxicity assays to compounds of interest. We extracted active and inactive compounds from the

screening results of each assay, then removed compounds with inconclusive or ambiguous results. We focused our analyses on two cytotoxicity assays: HEK293 and HepG2, for each of which at least 500 compounds are available for model training (**Table S1**).

Extracting chemical-gene connections from ComptoxAI for model construction

ComptoxAI^{19, 20} is a comprehensive graph knowledge base that contains curated relationships between chemicals, genes, assays, and many other entities. It contains two types of relationships linking compound and gene nodes: physical binding (with protein product) and expression-alteration (up-regulation/down-regulation). We extracted both types for chemicals present in the Tox21 datasets. We also constructed a hybrid type by combining the two types of relationships. We used the extracted relationships to assemble binary compound-gene matrices as input feature profiles.

To perform dimensionality reduction, we implemented a ReliefF-based method—namely, MultiSURF²¹—for ranking genes by relevance to the assay outcome of interest. MultiSURF takes in a labeled feature dataset (assay outcome as the label), ranks all features based on differences among neighboring instances. Specifically, intraclass differences will contribute negatively to feature relevance while interclass differences will contribute positively. A benchmark study showed that MultiSURF outperformed other methods in detecting genotype-phenotype associations²¹. We selected the top 100, 200, 300, 400, and 500 ranked genes of each dataset to proceed with the following analysis. Model selection by training loss was carried out to identify the optimal gene feature space (**Table S2**).

Constructing VNN with selected gene features and Reactome pathway hierarchy

DTox is a VNN model embedded with the Reactome pathway hierarchy that comprises root biological processes, child-parent pathway relations, and gene-pathway annotations (downloaded in Aug 2019)²². Each pathway is embedded as a module consisting of hidden neurons. AIDox inherits the basic structure of DTox while modifying input layers to consist of the gene features selected by MultiSURF. Guided by gene-pathway annotations, the gene features are connected to modules representing lowest-level pathways. These modules are then connected to deeper layers based on child-parent pathway relations until root biological processes are reached. Finally, the root biological processes are connected to an output layer representing the assay outcome. To trim the network's scale and prevent overfitting, we made the root biological process a hyperparameter of AIDTox. We selected four processes ('gene expression', 'immune system', 'metabolism', and 'signal transduction') and all possible combinations among them for the tuning process (15 values in total). These four processes were selected due to their broad coverage and direct involvement in cellular mechanisms of toxicity.

AIDTox also inherits the hybrid loss function of DTox combining both root and auxiliary loss terms: $BCELoss(y_{root}, y) + \alpha \sum_p \beta_p BCELoss(y'_{p-auxi}, y) + \lambda \|W\|_2$. The first term, namely root loss, denotes the binary cross entropy of final output y_{root} . The second term, namely auxiliary loss, denotes the binary cross entropy of the auxiliary scalar y'_{p-auxi} from each pathway module, with factor α ($= 0.5$) balancing the root and auxiliary terms, and factor β_p (computed as the inverse of pathway count within the corresponding hidden layer) normalizing terms across hidden layers. The third term denotes L_2 regularization with coefficient λ ($= 1e^{-4}$). The incorporation of auxiliary loss terms can prevent gradients from

vanishing in the lower hierarchy and facilitates learning new patterns from individual pathways.

Learning optimal AIDTox model for cytotoxicity prediction and interpretation

AIDTox adopts the same training scheme as DTox, including a training/testing/validation split by ratio of 7:1:2, optimization using the Adam algorithm²³ with mini-batch size of 32, a learning rate of 0.001, early stopping criterion with “patience” of 20, and hyperparameter tuning by grid search. We evaluated the performance of the optimal AIDTox model on the held-out validation set (**Table S3**), and compared it with three existing models: (i.) the optimal DTox model, (ii.) an optimal QSAR model based on a random forest classifier (derived from tuning six hyperparameters with 2800 combinations in total, listed in **Table S4**), and (iii.) an optimal QSAR model based on a gradient boosting classifier (derived from tuning five hyperparameters with 3000 combinations in total, listed in **Table S4**). For the two QSAR models (ii. and iii.), we adopted 166-bit binary MACCS fingerprints to quantify structural properties of compounds, which covers most of the interesting physicochemical features for drug discovery²⁴. We also implemented DTox’s model interpretation framework to identify paths from the VNN that can explain compound cytotoxicity (**Table S5**).

RESULTS

AIDTox employs curated chemical-gene connections to construct VNN

AIDTox predicts and explains compound response to toxicity assays with a knowledge-guided VNN. The knowledge incorporated in AIDTox comprises chemical-gene connections from ComptoxAI, as well as gene-pathway annotations and child-parent pathway relationships from Reactome (**Figure 1A**). Connections within the VNN are constrained to gene-pathway connections (input to first hidden layer) and child-parent pathway relations (after first hidden layer). In this study, we focused on two cell viability assays measuring compound cytotoxicity in HEK293 and HepG2 cells. Both datasets contain 1,367 compounds with connections in ComptoxAI, including 432 cytotoxic and 935 non-cytotoxic compounds in HEK293 cells, and 293 cytotoxic and 1,074 non-cytotoxic compounds in HepG2 cells. Among these compounds, 617 (45%) are connected to 991 genes with physical binding evidence, 1,237 (90%) are connected to 8,723 genes with expression-alteration evidence, and 1,367 (100%) are connected to 8,735 genes with both types of evidence. In all three cases, the number of gene features is much greater than the number of compound samples. To prevent overfitting, we reduced feature dimensions by selecting the top 100-500 predictive genes during model construction. In general, as more predictive genes were incorporated, we observed an improvement in the resulting model performance (decline in training loss) until 400 (**Figure S1**). Therefore, we proceeded with models built using the top 400 predictive genes.

Chemical-gene binding connections result in the best performing models

Using held-out validation sets of HEK293 and HepG2 cell viability data, we first compared the performance of models derived from three distinct connection types. In both datasets, models derived from binding connections significantly outperform the other types (**Figure 1B&C; Table S2**), as we observed no overlaps between the 95% confidence intervals of

compared metrics (AUROC, balanced accuracy, and F1-score) except for AUROC on the HepG2 viability dataset. We then compared the performance of AIDTox models derived from binding connections against three other classification algorithms, including our previous DTox model (derived from predicted compound-target interactions) and two QSAR models (derived from quantified structural properties). In both datasets, the 95% confidence intervals of performance metrics are highly overlapped among compared algorithms (**Figure 1D&E; Table S3**). Therefore, AIDTox achieved the same level of predictive performance as these well-established classification algorithms.

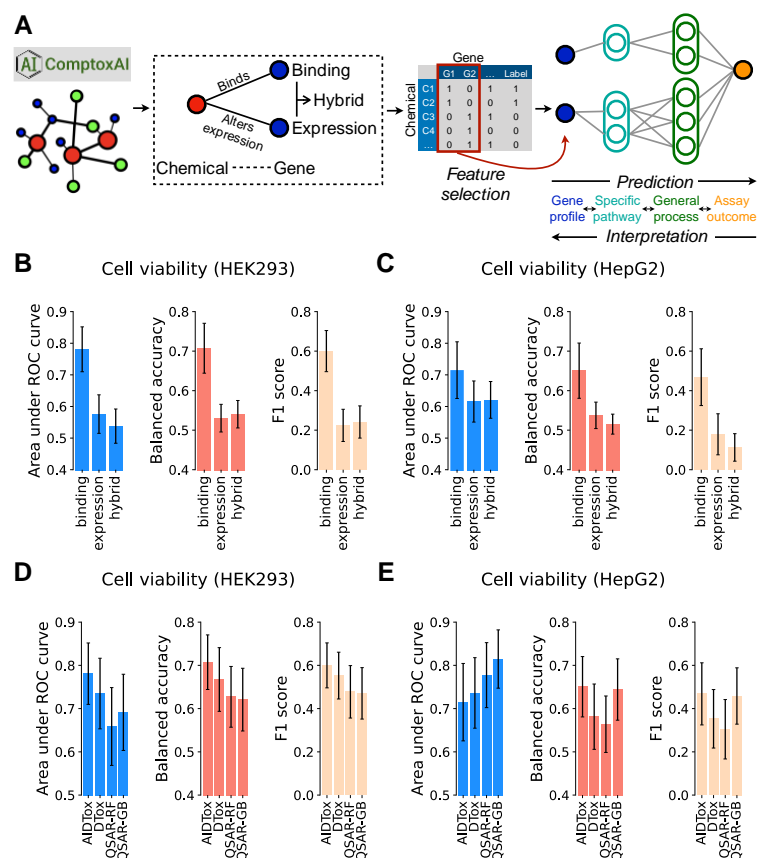


Figure 1 | Incorporating curated chemical-gene connections into VNN for toxicity prediction with AIDTox (A)

Three types of chemical-gene connections (binding, expression, and hybrid) are extracted from ComptoxAI to construct the input feature profile of AIDTox. Feature selection is implemented to identify the top gene features predictive of the outcome of interest. The selected profile is fed into a VNN, whose structure is guided by Reactome pathway hierarchy. Specific pathways and general processes are coded as modules by hidden neurons. **(B&C)** Barplots showing the comparison of validation performance across three connection types in two cell viability datasets: HEK293 **(B)** and HepG2 **(C)**.

Performance is measured by three metrics: area under ROC curve, balanced accuracy, and F1 score, with error bar showing the 95% confidence interval. **(D&E)** Barplots showing the comparison of validation performance across four models in two cell viability datasets: HEK293 **(D)** and HepG2 **(E)**. Three other models are considered: (i) our previous DTox model with inferred target profile as input, (ii) QSAR model by random forest with chemical fingerprint as input, and (iii) QSAR model by gradient boosting with chemical fingerprint as input.

AIDTox models benefit from a comprehensive gene feature space

A fundamental advantage of AIDTox over DTox comes from its enlarged gene feature space, providing an extensive profiling of the cellular activities of compounds. Overall, there is an increase from 361 genes in DTox to 991 genes in AIDTox, a 2.7-fold increase. Comparing the composition of gene features, we discovered that the increase in AIDTox is mainly driven by three target categories: ion channels, transporters, and others (**Figure 2A**). One example of the “others” category is the tubulin protein superfamily (**Figure 2C**), which polymerize into microtubules, a primary component of cytoskeleton involved in cell division. Due to the lack of binding data, tubulin proteins were not present in the DTox model. In ComptoxAI, the two families of tubulin proteins (α and β) are connected to antiparasitics (albendazole, and mebendazole) and antineoplastics (podofilox, vincristine sulfate, and docetaxel). Our AIDTox model further connects these drugs to HEK293 cytotoxicity via Rho GTPase signaling

by effectors such as formins and IQGAPs, which have been shown to regulate apoptosis in HEK293 cells²⁵.

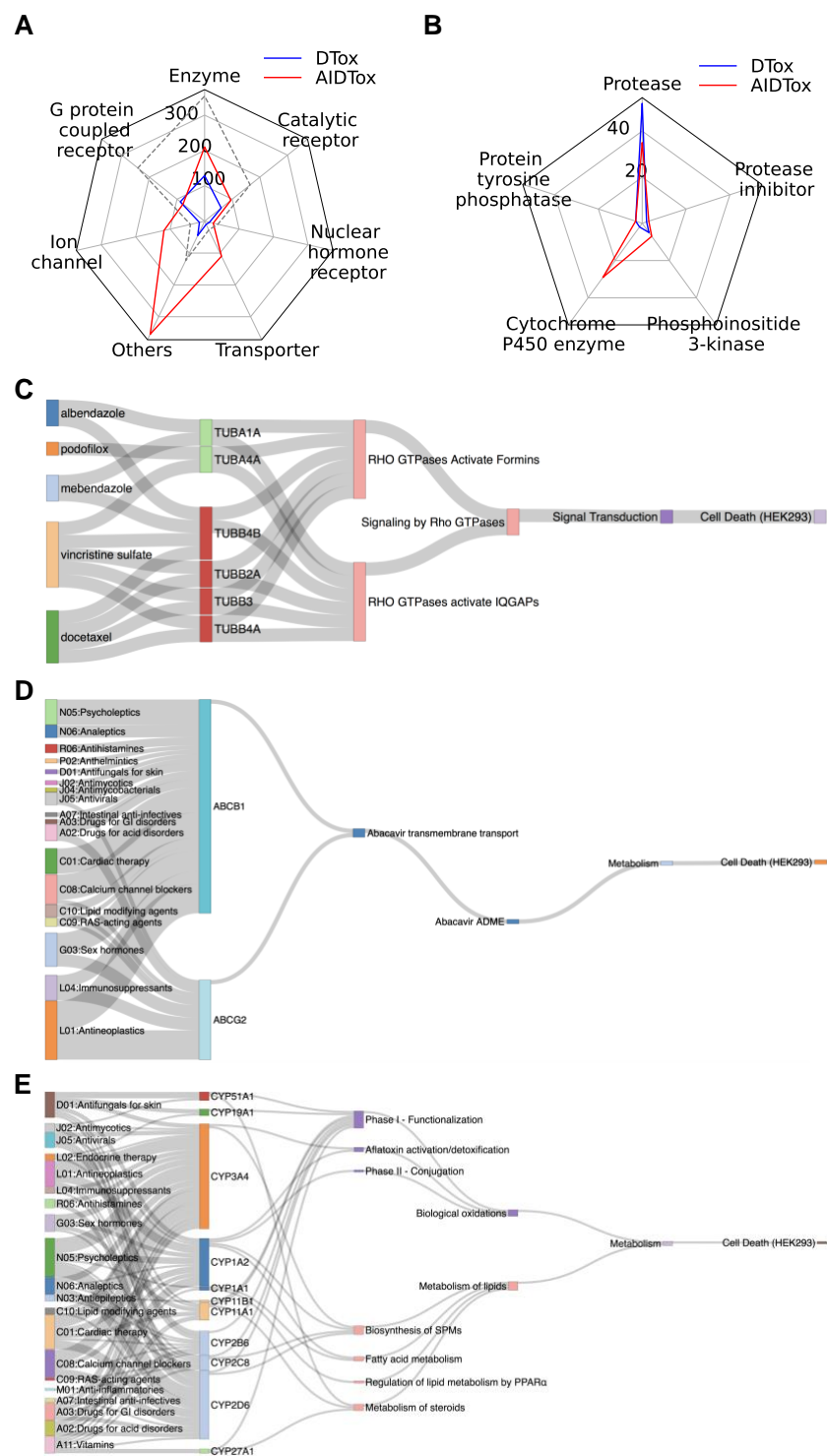


Figure 2 | Comprehensive explanation of HEK293 cytotoxicity with new features in AIDTox (A) Radar plot showing the comparison of gene category distributions among the features of DTox (blue solid line) and AIDTox (red solid line). The grey dashed line shows the hypothetical of a proportionate increase from DTox to AIDTox. **(B)** Radar plot showing the comparison of enzyme subcategory distributions among the features of DTox (blue solid line) and AIDTox (red solid line). **(C)** Sankey diagram showing the AIDTox explanation of HEK293 cytotoxicity for drugs targeting tubulin proteins. The paths (connecting drugs to HEK293 cell death) shown in the diagram are identified from the full network of VNN model by the AIDTox interpretation framework. Connections in the VNN are informed by ComptoxAI (chemical-gene) and Reactome (gene-pathway, child-parent pathway). Tubulin proteins are grouped and colored by the family. **(D)** Sankey diagram showing the AIDTox explanation of HEK293 cytotoxicity via ATP-binding cassette transporters (similar to C). Drugs are grouped and colored by the ATC subclass (first three digits). **(E)** Sankey diagram showing the AIDTox explanation of HEK293 cytotoxicity via cytochrome P450 enzymes (similar to C). Drugs are grouped and colored by the ATC subclass (first three digits). Cytochrome P450 enzymes are grouped and colored by the family. Metabolic pathways are grouped and colored by the general metabolic process they belong to (“Biological oxidations” or “Metabolism of lipids”).

New features in AIDTox are essential in drug metabolism and elimination processes
 We observed a disproportionate 11-fold increase (from 10 to 110) among transporters in the gene feature space of AIDTox. For instance, AIDTox connects drugs within nine ATC

classes (18 subclasses) to HEK293 cytotoxicity via two new features: transporters *ABCB1* and *ABCG2* (**Figure 2D**). These two ATP-binding cassette transporters are mainly responsible for pumping drugs out of the cell, thus they play a critical role in drug elimination. Therefore, inhibition of transporter-mediated elimination by drugs may prolong their cytotoxic effect on HEK293 cells²⁶.

While we did not observe a disproportionate increase among the general enzyme category, we did observe a 14.5-fold increase (from 2 to 29) among one subcategory: the cytochrome P450 enzymes (CYPs; **Figure 2B**). For instance, AIDTox connects drugs within nine ATC classes (20 subclasses) to HEK293 cytotoxicity via 20 CYPs of seven families (**Figure 2E**). These CYPs are the major enzymes involved in drug metabolism as they participate in metabolic pathways such as “biological oxidation” and “metabolism of lipids”. Meanwhile, CYP-mediated metabolic processes are the main source of reactive oxygen species, which can lead to cellular oxidative stress and trigger apoptosis/necrosis²⁷.

DISCUSSION

Rich knowledge in ComptoxAI provides an accurate and extensive profiling of chemical-gene connections. Here, we have explored the incorporation of these connections into knowledge-guided deep learning models for predicting and explaining compound cytotoxicity. We considered three types of connections for the task: physical binding, expression-alteration, and a hybrid type. Models derived from binding connections exhibit the best predictive performance, since physical binding is stronger evidence of direct interaction compared to expression alteration. In contrast to our previous work DTox, the new AIDTox model employs curated knowledge for generating input feature profile, thus is not prone to errors from binding prediction models. In addition, AIDTox is not restrained by the availability of compound-target binding data. The feature space of AIDTox comprises many genes with insufficient binding data, including ion channels, transporters, and CYP enzymes. These categories exhibit central roles in drug metabolism and elimination processes. Accordingly, they become an asset for prediction and explanation of complex toxicity outcomes, which may be triggered by multiple cellular mechanisms. For instance, AIDTox was able to connect dasatinib, a leukemia drug, to HEK293 cytotoxicity via multiple aspects of drug activity, including MAPK14-mediated regulation of apoptosis, CYP1A2/CYP3A4-mediated drug metabolism, and ABCB1/ABCG2-mediated drug elimination (**Figure S2**). It is also worth noticing the high interpretability of AIDTox is achieved without loss of accuracy. Therefore, we anticipate AIDTox to be applied in both prioritizing compounds for safety testing and generating new hypothesis for mechanistic investigation.

Despite these highlights, AIDTox did not significantly outperform existing methods in cytotoxicity prediction. This can be attributed to the sharp decline in sample size, as the profiling of chemical-gene connections by ComptoxAI only applies to well-studied compounds. A smaller sample size in turn leads to wider confidence intervals of model performance metrics. Consequently, even though we observed a moderate increase in the performance metrics, the increase is well within the 95% confidence interval. In the future, we expect the performance of AIDTox to be enhanced after chemical-gene connections

become available for more compounds. We also acknowledge the recent development of graph neural network-based link prediction algorithms, which can fill in missing connections for under-studied compounds. Such algorithms may help us increase the sample size for AIDTox training and enhance model performance. In addition to increasing sample size, we think the incorporation of context-specific knowledge, such as cell line-specific chemical-gene connections, may further enhance the performance of AIDTox.

FUNDING

This work was supported by NIH grants P30 ES013508, R01 LM010098, R01 AG066833, and K99 LM013646.

CONFLICTS OF INTERESTS

The authors declare no competing financial interest.

AUTHOR CONTRIBUTIONS

J.H.M. and Y.H. conceived the AIDTox project. J.H.M., Y.H., and J.D.R. designed the AIDTox model and data analysis workflow. Y.H. and J.D.R. performed the analysis. J.H.M., Y.H., and J.D.R. interpreted the results and wrote the paper. All authors read and approved the final manuscript.

REFERENCES

1. Hemmerich, J. and G.F. Ecker (2020). In silico toxicology: From structure–activity relationships towards deep learning and adverse outcome pathways. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 10, e1475.
2. Cherkasov, A., E.N. Muratov, D. Fourches, A. Varnek, Baskin, II, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, and A. Tropsha (2014). QSAR modeling: where have you been? Where are you going to? *J Med Chem* 57, 4977-5010.
3. Richard, A.M., R. Huang, S. Waidyanatha, P. Shinn, B.J. Collins, I. Thillainadarajah, C.M. Grulke, A.J. Williams, R.R. Lougee, R.S. Judson, K.A. Houck, M. Shobair, C. Yang, J.F. Rathman, A. Yasgar, S.C. Fitzpatrick, A. Simeonov, R.S. Thomas, K.M. Crofton, R.S. Paules, J.R. Bucher, C.P. Austin, R.J. Kavlock, and R.R. Tice (2021). The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. *Chem Res Toxicol* 34, 189-216.
4. Kleinstreuer, N.C., J. Yang, E.L. Berg, T.B. Knudsen, A.M. Richard, M.T. Martin, D.M. Reif, R.S. Judson, M. Polokoff, D.J. Dix, R.J. Kavlock, and K.A. Houck (2014).

- Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nat Biotechnol* 32, 583-91.
5. Sedykh, A., H. Zhu, H. Tang, L. Zhang, A. Richard, I. Rusyn, and A. Tropsha (2011). 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* 119, 364-70.
 6. Liu, J., K. Mansouri, R.S. Judson, M.T. Martin, H. Hong, M. Chen, X. Xu, R.S. Thomas, and I. Shah (2015). Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. *Chem Res Toxicol* 28, 738-51.
 7. Ammad-ud-din, M., E. Georgii, M. Gonen, T. Laitinen, O. Kallioniemi, K. Wennerberg, A. Poso, and S. Kaski (2014). Integrative and personalized QSAR analysis in cancer by kernelized Bayesian matrix factorization. *J Chem Inf Model* 54, 2347-59.
 8. Yamane, J., S. Aburatani, S. Imanishi, H. Akanuma, R. Nagano, T. Kato, H. Sone, S. Ohsako, and W. Fujibuchi (2016). Prediction of developmental chemical toxicity based on gene networks of human embryonic stem cells. *Nucleic Acids Res* 44, 5515-28.
 9. Capuzzi, S.J., R. Politi, O. Isayev, S. Farag, and A. Tropsha (2016). QSAR Modeling of Tox21 Challenge Stress Response and Nuclear Receptor Signaling Toxicity Assays. *Frontiers in Environmental Science* 4.
 10. Zhang, J., D. Mucs, U. Norinder, and F. Svensson (2019). LightGBM: An Effective and Scalable Algorithm for Prediction of Chemical Toxicity-Application to the Tox21 and Mutagenicity Data Sets. *J Chem Inf Model* 59, 4150-4158.
 11. Mayr, A., G. Klambauer, T. Unterthiner, and S. Hochreiter (2016). DeepTox: Toxicity Prediction using Deep Learning. *Frontiers in Environmental Science* 3.
 12. Idakwo, G., S. Thangapandian, J.t. Luttrell, Z. Zhou, C. Zhang, and P. Gong (2019). Deep Learning-Based Structure-Activity Relationship Modeling for Multi-Category Toxicity Classification: A Case Study of 10K Tox21 Chemicals With High-Throughput Cell-Based Androgen Receptor Bioassay Data. *Front Physiol* 10, 1044.
 13. Matsuzaka, Y. and Y. Uesawa (2020). Molecular Image-Based Prediction Models of Nuclear Receptor Agonists and Antagonists Using the DeepSnap-Deep Learning Approach with the Tox21 10K Library. *Molecules* 25.
 14. Polishchuk, P. (2017). Interpretation of Quantitative Structure-Activity Relationship Models: Past, Present, and Future. *J Chem Inf Model* 57, 2618-2639.
 15. Ma, J., M.K. Yu, S. Fong, K. Ono, E. Sage, B. Demchak, R. Sharan, and T. Ideker (2018). Using deep learning to model the hierarchical structure and function of a cell. *Nat Methods* 15, 290-298.
 16. Elmarakeby, H.A., J. Hwang, R. Arafeh, J. Crowdis, S. Gang, D. Liu, S.H. AlDubayan, K. Salari, S. Kregel, C. Richter, T.E. Arnoff, J. Park, W.C. Hahn, and E.M. Van Allen (2021). Biologically informed deep neural network for prostate cancer discovery. *Nature* 598, 348-352.
 17. Lin, C.H. and O. Lichtarge (2021). Using Interpretable Deep Learning to Model Cancer Dependencies. *Bioinformatics*.
 18. Hao, Y., J.D. Romano, and J.H. Moore (2022). Knowledge-guided deep learning models of drug toxicity improve interpretation. *Patterns* 3, 100565.
 19. Romano, J.D., Y. Hao, J.H. Moore, and T.M. Penning (2022). Automating Predictive Toxicology Using ComptoxAI. *Chemical Research in Toxicology*.

20. Romano, J.D., Y. Hao, and J.H. Moore. *Improving QSAR Modeling for Predictive Toxicology using Publicly Aggregated Semantic Graph Data and Graph Neural Networks*. in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022*. 2021. World Scientific.
21. Urbanowicz, R.J., R.S. Olson, P. Schmitt, M. Meeker, and J.H. Moore (2018). Benchmarking relief-based feature selection methods for bioinformatics data mining. *J Biomed Inform* 85, 168-188.
22. Jassal, B., L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio (2020). The reactome pathway knowledgebase. *Nucleic Acids Res* 48, D498-D503.
23. Kingma, D.P. and J. Ba (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
24. Cereto-Massague, A., M.J. Ojeda, C. Valls, M. Mulero, S. Garcia-Vallve, and G. Pujadas (2015). Molecular fingerprint similarity search in virtual screening. *Methods* 71, 58-63.
25. Quinn, N.P., L. García-Gutiérrez, C. Doherty, A. von Kriegsheim, E. Fallahi, D.B. Sacks, and D. Matallanas (2021). IQGAP1 is a scaffold of the core proteins of the hippo pathway and negatively regulates the pro-apoptotic signal mediated by this pathway. *Cells* 10, 478.
26. Eadie, L., T. Hughes, and D. White (2014). Interaction of the efflux transporters ABCB1 and ABCG2 with imatinib, nilotinib, and dasatinib. *Clinical Pharmacology & Therapeutics* 95, 294-306.
27. Veith, A. and B. Moorthy (2018). Role of cytochrome P450s in the generation and metabolism of reactive oxygen species. *Current opinion in toxicology* 7, 44-51.