National Institute of Environmental Health Sciences
Division of Translational Toxicology

Applying In Silico Toxicity Models Across the Tox21 Chemical Universe

S. Lee[1], A. Borrel[2], K. To[2], T. Li[3], Z. Liu[3], E. Benfenati[4] A. Roncaglioni[4], A. Manganaro[5], K. Mansouri[6], N. Kleinstreuer[6], W. Tong[3]

[1]Duke University, Durham, NC, USA; [2]Inotiv, RTP, NC, USA; [3]FDA/NCTR, Jefferson, AR, USA; [4]Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy; [5]KODE srl, Pisa, Italy; and [6]NIH/NIEHS/DTT/NICEATM, RTP, NC, USA

Introduction

- Guideline rodent carcinogenicity studies[1] generally require about 500 rodents, cost an average of $1.1 million, and generate results of questionable relevance to humans.
- Drug-induced liver injury (DILI) is a major cause of failure for new drugs in clinical trials and is poorly predicted by rodent studies.
- New approach methodologies (NAMs) are available that use in silico and in vitro methods to predict carcinogenicity and DILI.
- Quantitative structure–activity relationship (QSAR) approaches have been developed that can identify potentially carcinogenic and hepatotoxic chemicals. These methods can provide insights into bioactivity of novel chemicals and may have potential regulatory applications given appropriate validation.

Objectives

We applied three QSAR models to predict carcinogenicity and hepatotoxicity of chemicals in the Tox21 chemical set:

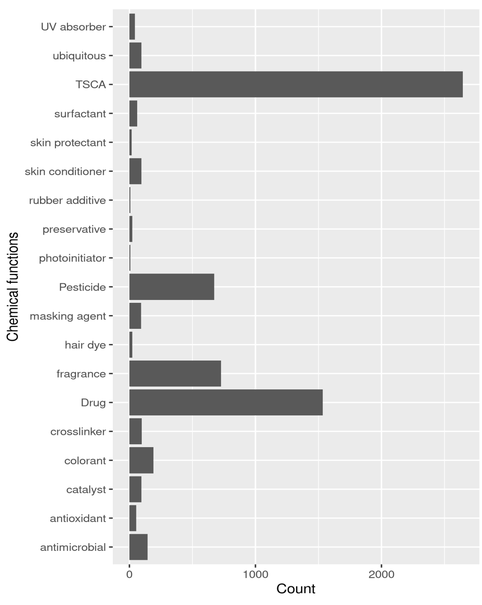
- **DeepCarc**, carcinogenicity model developed by the National Center for Toxicological Research (NCTR), U.S. Food and Drug Administration (FDA).
- **JANUS** (Joining environmental, ecotoxicological and toxicological Assessment of chemical substances with Non-testing methods within a Unified Screening), carcinogenicity model developed by the Istituto di Ricerche Farmacologiche Mario Negri.
- **DeepDILI**, hepatotoxicity model developed by FDA NCTR.

This presentation:

- Compares the predictions of the two carcinogenicity models (JANUS and DeepCarc) to the set of known carcinogens and discusses model limitations.
- Identifies chemicals that are predicted to be carcinogens with high probability, including those in the Tox21 set.
- Identifies chemicals from Tox21 that are predicted highly likely to induce DILI.

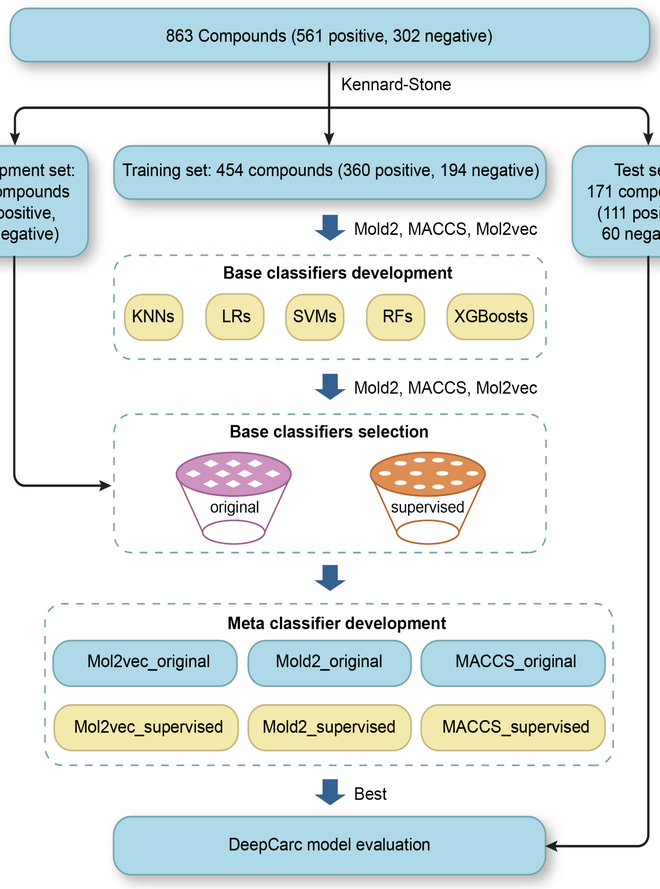
1. Tox21 Chemical Universe

The Tox21 library includes around 10,000 chemicals (8,305 unique structures)[2]. This diverse set of chemicals was used in the U.S. federal research collaboration focused on developing methods to evaluate the safety of commercial chemicals, pesticides, food additives/contaminants, and medical products.



The bar graph shows categories of 4,950 Tox21 chemicals included in the U.S. Environmental Protection Agency (EPA) Consumer Products Database [3]. TSCA, Toxic Substances Control Act (EPA).

2. DeepCarc Carcinogenicity Model



The training set was built from chemicals in the NCTR's Liver Cancer Database[4] and included mostly liver carcinogens.

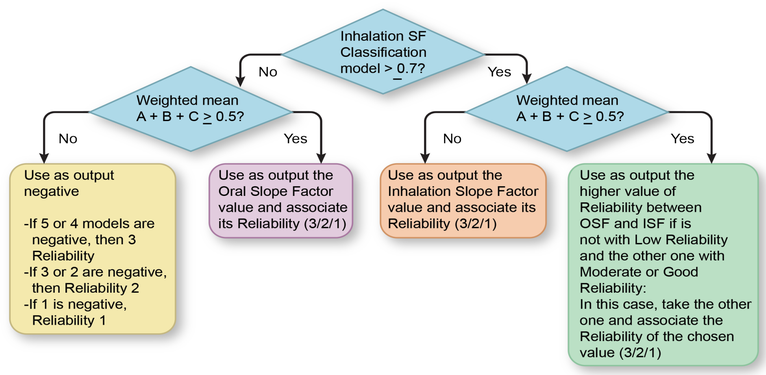
DeepCarc[5] was developed using five conventional machine learning algorithms and a neural network to generate a probability that a chemical is a carcinogen.

3. JANUS Carcinogenicity Model

The JANUS carcinogenicity model[6] is a consensus model implemented in a decision tree. It includes both classification machine learning-based models and chemical substructure alert searches. It is available on VegaHub [6].

The JANUS training data included:

- Chemicals in the Carcinogenic Potency Database with animal bioassay data (805 chemicals)[7]
- ANTARES carcinogenicity data set (1,543 chemicals)[8]
- 986 rodent carcinogens[9]



4. DeepDILI - Hepatoxicity Model

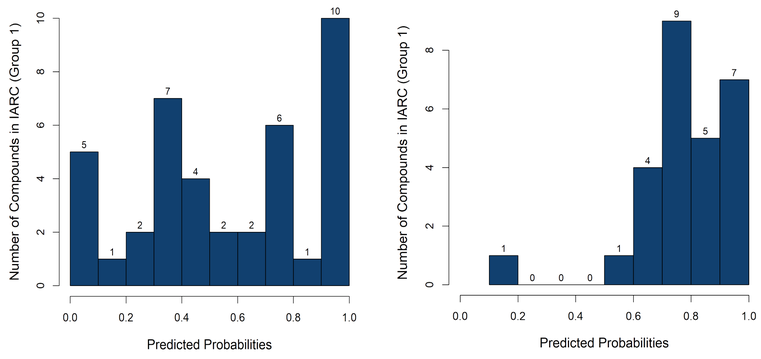
DeepDILI[10] predicts chemicals that can induce liver injury. It was developed using a similar approach as DeepCarc via a combination of machine learning approaches in a neural network. The training set was 1,002 drugs extracted from DrugBank[11] and FDAlabel[12] databases.

Chemicals	Count
Drug inducing liver injury (active)	604
Drug not inducing liver injury (inactive)	398
Total	**1,002 (753 in training)**

5. DeepCarc / JANUS on Carcinogens

- The IARC Monographs program[13] evaluates chemicals that can cause cancer in humans. The International Agency for Research on Cancer (IARC) has evaluated more than 1000 agents and identified 122 human carcinogens ("Class 1").
- For this project we used 41 Class 1 carcinogens that have well-defined QSAR-ready structures.
- We explored whether DeepCarc and JANUS could correctly identify the 41 IARC Class 1 carcinogens as carcinogens.

DeepCarc Predictions ***JANUS Predictions***



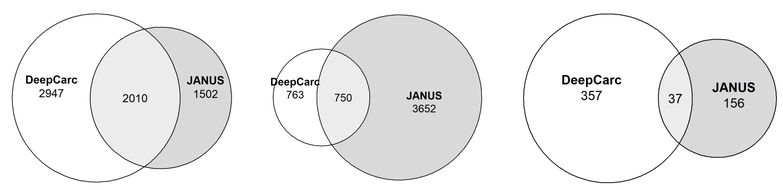
In the histograms above, chemicals assigned a probability of greater than 0.5 are considered to be predicted carcinogens. DeepCarc correctly predicted 21 of 41 chemicals, while JANUS correctly predicted 26 of 27 chemicals.

This difference can be explained by the data sets used to build the models. While some of the same chemicals were used in the training sets for both models, DeepCarc focused on liver carcinogens while JANUS was developed more broadly from a consensus modeling of a more diverse set of chemicals.

6. Tox21 Carcinogenicity Predictions

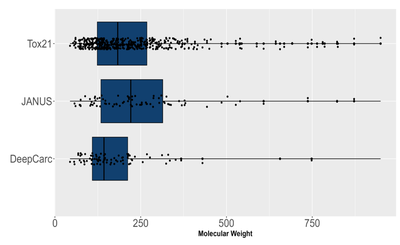
Risk Category	Low-Risk (PC < 0.2, PJ < 0.2)	Carcinogen Risk (PC > 0.5, PJ > 0.5)	High-Risk (PC > 0.8, PJ > 0.8)
DeepCarc (PC)	57.6% (4,957)	9.7% (838)	4.6% (394)
JANUS (PJ)	44.4% (3,512)	55.6% (4,402)	2.4% (193)

Overlap Between the Predicted Chemicals



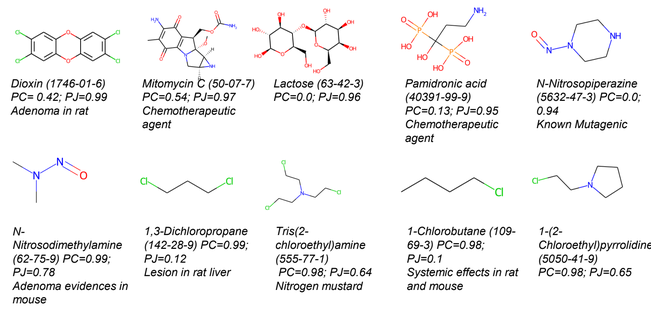
We then used DeepCarc and JANUS to identify chemicals in the Tox21 library that might be carcinogens. Overall JANUS predicted more chemicals to be carcinogenic (probability > 0.5) than DeepCarc. However, DeepCarc predicted more high-risk carcinogens (probability > 0.8).

7. Chemicals Predicted to be Carcinogens



A structural analysis of chemicals predicted by DeepCarc to be carcinogenic with probability >0.8 showed that these have a significantly lower molecular weight (<150 Da) than the average molecular weight of all chemicals in Tox21.

We show below the structures of some chemicals that DeepCarc and JANUS predicted highly likely to be carcinogens with some evidence supporting their carcinogenic potential.



Dioxin (1746-01-6) PC= 0.42; PJ=0.99 Adenoma in rat

Mitomycin C (50-07-7) PC=0.54; PJ=0.97 Chemotherapeutic agent

Lactose (63-42-3) PC=0.0; PJ=0.96

Pamidronic acid (40391-99-9) PC=0.13; PJ=0.95 Chemotherapeutic agent

N-Nitrosopiperazine (5632-47-3) PC=0.0; 0.94 Known Mutagenic

N-Nitrosodimethylamine (62-75-9) PC=0.99; PJ=0.78 Adenoma evidences in mouse

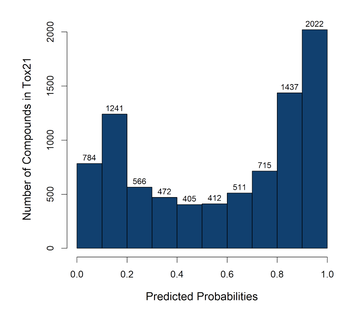
1,3-Dichloropropane (142-28-9) PC=0.99; PJ=0.12 Lesion in rat liver

Tris(2-chloroethyl)amine (555-77-1) PC=0.98; PJ=0.64 Nitrogen mustard

1-Chlorobutane (109-69-3) PC=0.98; PJ=0.1 Systemic effects in rat and mouse

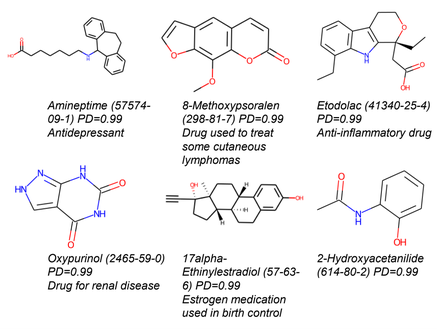
1-(2-Chloroethyl)pyrrolidine (5050-41-9) PC=0.98; PJ=0.65

8. Tox21 Hepatoxicity Predictions



We used DeepDILI to identify chemicals in the Tox21 library that might be hepatotoxins. Over 2,000 chemicals are predicted with a high probability (>0.9) to induce liver injury. Most of these are drugs, see below.

Risk Category	Percentage
Low-Risk (P < 0.1)	9.1% (784)
DILI Risk (P ≥ 0.5)	58.8% (4,223)
High-Risk (P ≥ 0.9)	23.5% (2,022)



Amineptime (57574-09-1) PD=0.99 Antidepressant

8-Methoxypsoralen (298-81-7) PD=0.99 Drug used to treat some cutaneous lymphomas

Etodolac (41340-25-4) PD=0.99 Anti-inflammatory drug

Oxypurinol (2465-59-0) PD=0.99 Drug for renal disease

17alpha-Ethinylestradiol (57-63-6) PD=0.99 Estrogen medication used in birth control

2-Hydroxyacetanilide (614-80-2) PD=0.99

9. Hepatoxic Chemicals

We used ChemMaps.com[14] to compare the properties of chemicals predicted to induce DILI with other chemical sets: approved drugs (green in above graphic), high-risk chemicals (gray), and withdrawn drugs (purple). The overlap between the Tox21 chemicals predicted by DeepDILI to be potential hepatotoxins (PD > 0.8) with the drug chemical space shows that many drugs should still be considered for DILI as a side effect.

Conclusions

- We applied DeepCarc and JANUS to the Tox21 chemical set and found that ~5% of the Tox21 chemicals are predicted with high probability by both models to be carcinogens.
- The comparison of these carcinogenicity models showed that DeepCarc performs better at predicting liver carcinogens, while JANUS predicts carcinogenicity more broadly.
- When we applied DeepDILI to the Tox21 chemical set we found that 23.5% of the chemicals were predicted to have a high probability of inducing DILI; most are drugs.
- The overlap with the drug space and the highly predicted DILI shown that many approved drugs could induce DILI.

References and Acknowledgments

References

1. https://www.ema.europa.eu/en/ich-s1a-need-carcinogenicity-studies-pharmaceuticals-scientific-guideline
2. Chem Res Toxicol 2021; 34:189−216. DOI: 10.1021/acs.chemrestox.0c00264
3. https://www.epa.gov/chemical-research/chemical-and-products-database-cpdat
4. https://www.foodrisk.org/resources/display/96
5. Front Artif Intell 2021 Nov18; 4:757780. DOI: 10.3389/frai.2021.757780
6. https://www.vegahub.eu/janus-the-tools-for-prioritization-and-screening-of-chemical-substances-freely-available/
7. https://www.nlm.nih.gov/databases/download/cpdb.html
8. Chem Cent J 2010 Jul29; 4 (Suppl 1):S3. DOI: 10.1186/1752-153X-4-S1-S3
9. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 2016 Apr2; 34(2):97–113. DOI: 10.1080/10590501.2016.1166879
10. Chem. Res. Toxicol 2021; 34:550−565. DOI: 10.1021/acs.chemrestox.0c00374
11. https://go.drugbank.com/
12. https://www.fda.gov/science-research/bioinformatics-tools/fdalabel-full-text-search-drug-product-labeling
13. https://monographs.iarc.who.int/
14. https://sandbox.ntp.niehs.nih.gov/chemmaps/

Acknowledgments

This project was funded in whole or in part with federal funds from the NIEHS, NIH under Contract No. HHSN273201500010C. The authors declare there exists no conflict of interest.
Since the poster was written as part of the official duties of the authors, it can be freely copied.