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], W. Tong[3], E. Benfenati[4], A. Roncaglioni[4], A. Manganaro[5], K. Mansouri[6], N. Kleinstreuer[6]

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

Abstract Number 3740
Poster Number P226

Introduction

- Guideline rodent carcinogenicity studies[1] generally require ~500 rodents, cost an average of $1.1 million, and frequently 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 are now available that use in silico/in vitro methods to predict carcinogenicity and DILI.
- Quantitative structure–activity relationship (QSAR) approaches have successfully been developed and applied to 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:

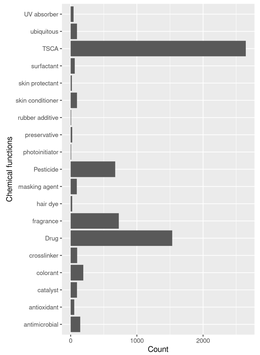
- **DeepCarc**, carcinogenicity model developed by the National Center for Toxicological Research, U.S. Food and Drug Administration (FDA).
- **JANUS** (Joining environmental, ecotoxicological and toxicological Assessment of chemical substances with Non-testing methods within a Unified Screening), carcinogenicity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

This presentation:

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

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.

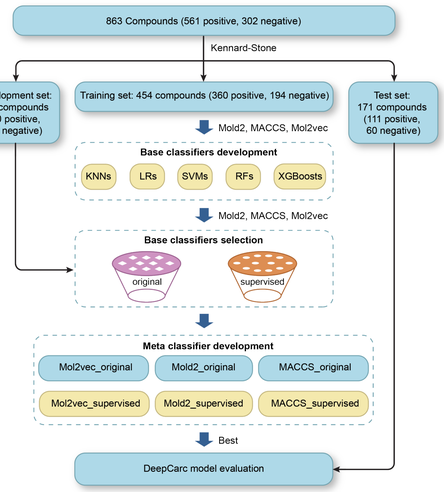


A representation of the 4,950 Tox21 chemicals included in the Consumer Products Database[3] shows that chemicals are associated with diverse classes of use including drugs, pesticides, and consumer products.

Carcinogen Class 1 from IARC monograph

- The IARC Monographs program[4] evaluates chemicals that can cause cancer in humans. They have evaluated more than 1000 agents and identified 122 carcinogens (Class 1).
- For this project we used 41 Class 1 carcinogens that have well-defined QSAR-ready structures.

The DeepCarc Carcinogenicity Model



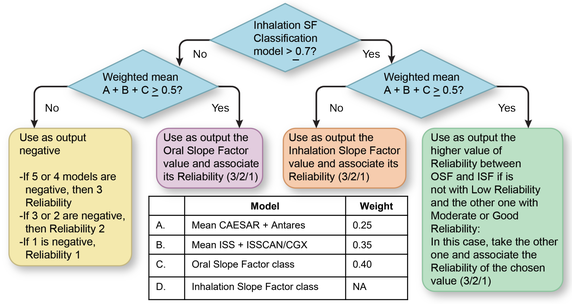
DeepCarc[5] was developed using five conventional machine learning algorithms and a neural network to generate a probability that a chemical is a carcinogen.

The training set was built from chemicals in the National Center for Toxicological Research's Liver Cancer Database[6] and included mostly liver carcinogens.

JANUS Carcinogenicity Model

The JANUS carcinogenicity model[7] is a consensus model implemented in a decision tree. It includes classification machine learning based models as well as chemical substructure alert searches. It is available on Vega Hub (https://www.vegahub.eu/portfolio-item/janus/).

The JANUS training data included various data sets:
- 805 chemicals from the Carcinogenic Potency Database (CPDB) of animal cancer bioassay data[8]
- ANTARES carcinogenicity dataset including 1,543 chemicals tested on rat[9]
- 986 rodent carcinogens from the ISSCAN data set and the Kirkland et al. 2005 publication[10-11]



	Model	Weight
A.	Mean CAESAR + Antares	0.25
B.	Mean ISS + ISSCAN/CGX	0.35
C.	Oral Slope Factor class	0.40
D.	Inhalation Slope Factor class	NA

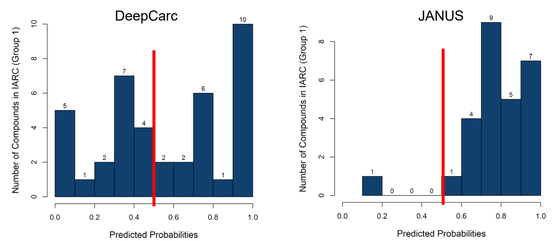
DeepDILI Model

DeepDILI[12] is used to predict if a chemical 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 development set is 1,002 drugs extracted from DrugBank and FDAlabel databases.

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

DeepCarc and JANUS Predict Known Carcinogens

Among the chemicals identified as Class 1 carcinogens by the International Agency for Research on Cancer, 41 have well-characterized structures. We explored whether DeepCarc and JANUS could correctly identify these chemicals as carcinogens.



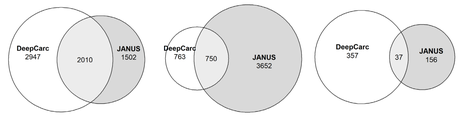
The bargraphs show that the JANUS model performed better on this set of chemicals. If chemicals assigned a probability of >0.5 are considered to be predicted carcinogens, then DeepCarc correctly predicted 21 of 41 chemicals, while JANUS correctly predicted 40 of 41 chemicals. There is no correlation between carcinogenic probabilities predicted using JANUS and using DeepCarc.

This difference can be explained by the dataset used to build both models. DeepCarc is focused on liver carcinogens while JANUS was developed more broadly from consensus modeling of a more diverse set of chemicals. Some of the chemicals are also used in the training set of JANUS models.

Tox21 Carcinogenicity Predictions

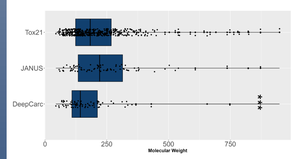
Risk Category	Low-Risk (PC < 0.2, PJ < 0.2)	Carcinogen Risk (PC > 0.5, PJ > 0.5)	High-Risk (PC > 0.8, PJ > 0.8)
DeepCarc (PC)	57.6% (4,957)	9.7% (838)	4.6% (394)
JANUS (PJ)	44.4% (3,512)	55.6% (4,402)	2.4% (193)

Overlap between the predicted chemicals



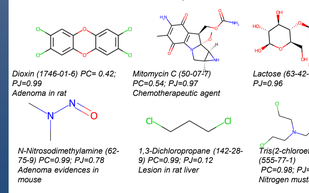
We then used DeepCarc and JANUS to identify chemicals in the Tox21 library that might be carcinogens. Overall JANUS predicts more carcinogenic chemicals with a probability > 0.5 than DeepCarc. However, DeepCarc predicts more high-risk carcinogens (probability > 0.8).

Tox21 Chemicals Highly Predicted to be Carcinogens



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

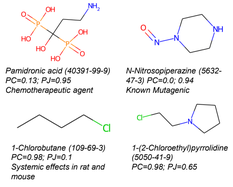
We show below the structures of some chemicals that DeepCarc and JANUS considered highly likely to be carcinogens with some evidence supporting their carcinogenic potential.



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

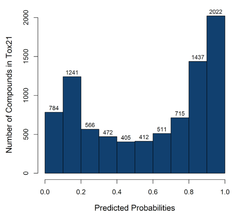
1,3-Dichloropropane (142-28-9) PC=0.99; PJ=0.12
Lesion in rat liver

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

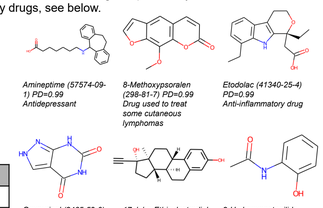
1-(2-Chloroethyl)pyrrolidine (5050-41-9)
PC=0.98; PJ=0.65

Tox21 Predictions by DeepDILI



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.

The chemicals with the highest probability to induce DILI included mostly drugs, see below.



Amineptine (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

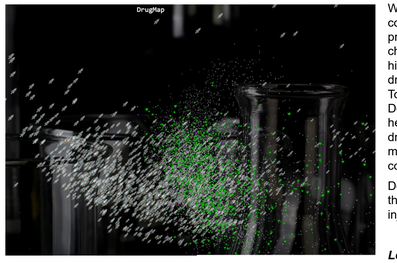
Oxypurinol (2465-59-0) PD=0.99
Drug for renal disease

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

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

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)

Chemicals Highly predicted DILI and the Drug Space



Overlap between highly predicted DILI chemicals and the drug chemicals space included in chemmaps.com[13]. (Approved drugs in green, withdrawn drugs in purple, and high-risk compounds as gray rockets)

We used ChemMaps.com to compare the properties of chemicals predicted to induce DILI with other chemical sets: approved drugs, high-risk chemicals, and withdrawn drugs. 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.

DeepDILI can help to predict drugs that are the most likely to cause liver injury.

Learn more about ChemMaps.com: Borrel et al., Abstract 3742/Poster P228

Conclusion

- We applied DeepCarc and JANUS to the Tox21 chemical set and found that about 5% of the Tox21 chemicals are predicted with high probability 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

1. https://www.ema.europa.eu/en/ich-s1a-need-carcinogenicity-studies-pharmaceuticals-scientific-guideline
2. Chem. Res. Toxicol.2021, 34, 189−216, 10.1021/acs.chemrestox.0c00264
3. https://www.epa.gov/chemical-research/chemical-and-products-database-cpdat
4. https://monographs.iarc.who.int/
5. Front. Artif. Intell. 4:757780. 10.3389/frai.2021.757780
6. https://www.foodrisk.org/resources/display/96
7. https://www.vegahub.eu/janus-the-tools-for-prioritization-and-screening-of-chemical-substances-freely-available/
8. https://www.ntm.nih.gov/databases/download/cpdb.html
9. Chemistry Central Journal 4 (Suppl 1), S3 (2010). 10.1186/1752-153X-4-S1-S3
10. https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=186943&Lab=NCCT
11. Br J Pharmacol. 2011 Mar; 162(6): 1250–1258. 10.1111/j.1476-5381.2010.01131.x
12. Chem. Res. Toxicol.2021, 34, 550−565. 10.1021/acs.chemrestox.0c00374
13. https://sandbox.ntp.niehs.nih.gov/chemmaps/

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.