Generalised Read-Across (GenRA) Prediction: Updating the Dashboard implementation and using genra-py as a standalone application

Abstract 232

I Shah[1], T Tate[1], L Girardin[2], S Hamilton[2], B Sharma[1], TN Brown[3], K Baron Furuyama[1], P Do[1], G Patlewicz[1*]
[1] Center for Computational Toxicology & Exposure (CCTE), U.S. Environmental Protection Agency, RTP, NC; [2]GDIT, Falls Church, VA; [3]CCTE, U.S. EPA, Duluth, MN

www.epa.gov

OBJECTIVES

- GenRA is a data-driven approach to estimate properties of chemicals from similar analogues.
- GenRA attempts to mimic the expert driven read-across workflow.
- We provide an update to the web application implemented as part of the EPA CompTox Chemicals Dashboard and present a new python package called genra-py that has been created that can be freely used for chemical safety analysis and risk assessment applications

APPROACH

- The web application implemented is targeted towards end-users already familiar with the read-across concept
- The web application is an update of the existing GenRA with refreshed toxicity and chemistry datasets
- The python package called genra-py is targeted for data scientists familiar with the python language who wish to use GenRA for many chemicals but with their own datasets.

MAIN RESULTS

- Look and feel of the User Interface is similar but with enhancements to improve user experience and refreshed datasets specifically ToxRefDB v2
- Genra-py is freely available from https://github.com/i-shah/genra-py under MIT license. Jupyter notebooks which document examples of previous analysis as well as formating of user input files are provided

IMPACT/SIGNIFICANCE

- Two implementations of GenRA are now available – an updated web application with the latest underlying chemistry and toxicity data as well as a standalone python package – genra-py.
- We anticipate these versions will facilitate new features and functionalities commensurate with our recent research findings to be more efficiently implemented. See posters 231 and 245 by W Jenkins and T Tate.

Disclaimer: The views expressed are those of the authors and do not necessarily reflect the views or policies of the US EPA.

Generalised Read-Across (GenRA) Prediction: Updating the Dashboard implementation and using genra-py as a standalone application

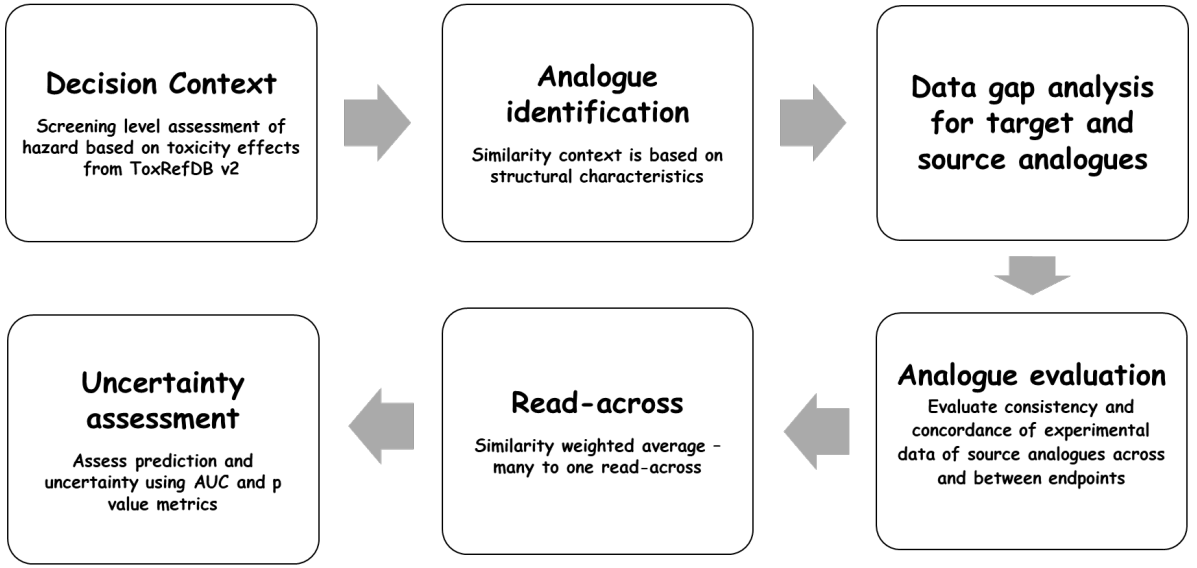
Abstract 232

I Shah[1], T Tate[1], L Girardin[2], S Hamilton[2], B Sharma[1], TN Brown[3], K Baron Furuyama[1], P Do[1], G Patlewicz[1*]
[1] Center for Computational Toxicology & Exposure (CCTE), U.S. Environmental Protection Agency, RTP, NC; [2]GDIT, Falls Church, VA; [3]CCTE, U.S. EPA, Duluth, MN

www.epa.gov

OBJECTIVES

- GenRA is a data-driven approach to estimate properties of chemicals from similar analogues.
- GenRA attempts to mimic the expert driven read-across workflow (Figure 1).
- We provide an update to the web application implemented as part of the EPA CompTox Chemicals Dashboard and present a new python package called genra-py that has been created that can be freely used for chemical safety analysis and risk assessment applications



- Figure 1: Read-across workflow implemented in the GenRA web interface

Generalised Read-Across (GenRA) Prediction: Updating the Dashboard implementation and using genra-py as a standalone application

Abstract 232

I Shah[1], T Tate[1], L Girardin[2], S Hamilton[2], B Sharma[1], TN Brown[3], K Baron Furuyama[1], P Do[1], G Patlewicz[1*]
[1] Center for Computational Toxicology & Exposure (CCTE), U.S. Environmental Protection Agency, RTP, NC; [2]GDIT, Falls Church, VA; [3]CCTE, U.S. EPA, Duluth, MN

www.epa.gov

MAIN RESULTS: Web application

Targeted towards end-users familiar with the read-across concept

- **Key updates:**
- **Decoupled GenRA from the EPA CompTox Chemicals Dashboard – entry point is direct to the GenRA User Interface to allow search for chemicals by Chemical Name, DSSTox Substance Identifier (DTXSID) or CASRN**
- **MongoDB collections updated to include all compounds registered in the latest snapshot of DSSTox**
- **Chemical fingerprint collections updated for Morgan, Torsion and Chemotype (ToxPrint) fingerprints**
- **Tox21 fingerprint collection sunsetted and ToxCast fingerprints now incorporate both ToxCast and Tox21 data**
- **Updated underlying toxicity data - migrated from Toxicity Reference Database v1 (ToxRefDB v1) to ToxRefDB v2 – restructured the data based on the schema changes in ToxRefDB v2**
- **Data underpinning GenRA no longer managed based on primary source databases to ensure seamless updates**
- **Code for GenRA refactored, re-written from Python 2 to Python 3. All API services rewritten.**
- **Look and feel of the User Interface is similar but with enhancements to improve user experience.**

Generalised Read-Across (GenRA) Prediction: Updating the Dashboard implementation and using genra-py as a standalone application

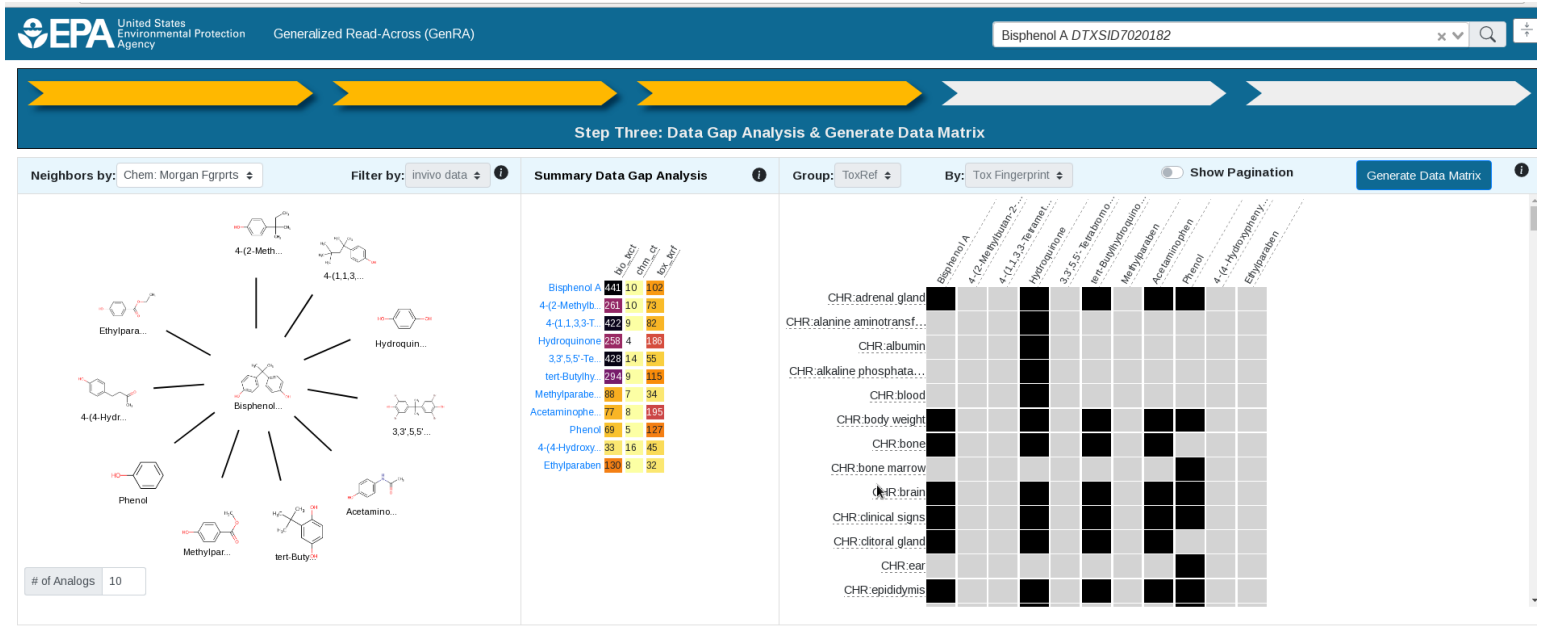
Abstract 232

I Shah[1], T Tate[1], L Girardin[2], S Hamilton[2], B Sharma[1], TN Brown[3], K Baron Furuyama[1], P Do[1], G Patlewicz[1*]

[1] Center for Computational Toxicology & Exposure (CCTE), U.S. Environmental Protection Agency, RTP, NC; [2]GDIT, Falls Church, VA; [3]CCTE, U.S. EPA, Duluth, MN

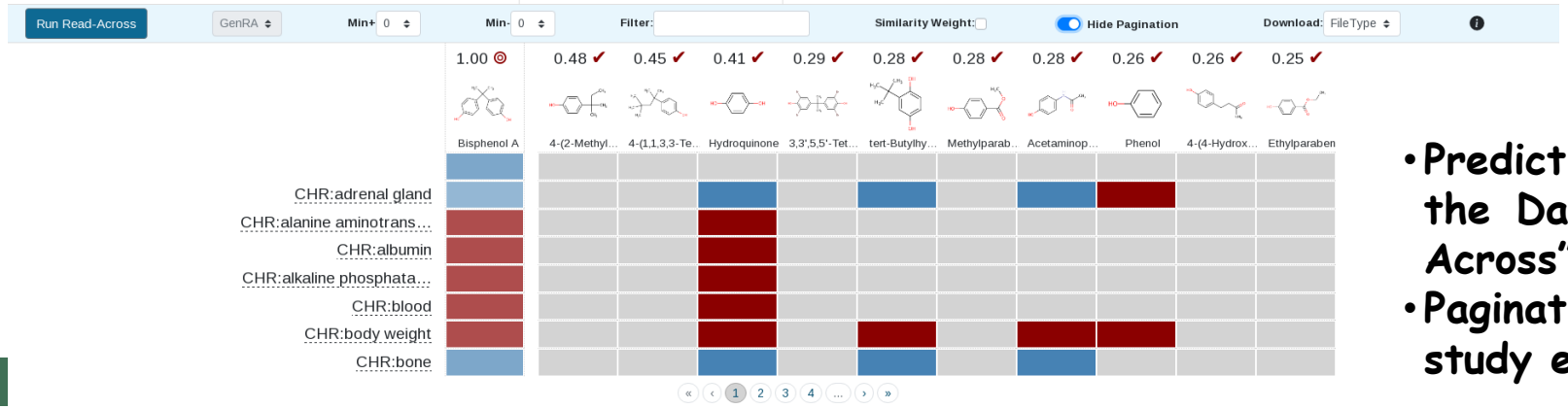
www.epa.gov

MAIN RESULTS: Web application



role	target	analog
preferred name	Bisphenol A	4-(2-Methylbutan-2-yl)phenol
dsstox_sid	DTXSID7020182	DTXSID8021771
molecular weight	228.291	164.248
similarity	1	0.483870968
	GenRA TN Act=0 (0) AUC=0 p=1	no_data
CHR:adrenal gland	GenRA TN Act=0 (0.212) AUC=0 p=0.8	no_data
CHR:alanine aminotransferase (alt/sgpt)	GenRA Pos Act=1 (1) AUC=0 p=1	no_data
CHR:albumin	GenRA Pos Act=1 (1) AUC=0 p=1	no_data
CHR:alkaline phosphatase (alp/alk)	GenRA Pos Act=1 (1) AUC=0 p=1	no_data
CHR:blood	GenRA Pos Act=1 (1) AUC=0 p=1	no_data
CHR:body weight	GenRA TP Act=1 (1) AUC=0 p=1	no_data
CHR:bone	GenRA TN Act=0 (0) AUC=0 p=1	no_data
CHR:bone marrow	GenRA Pos Act=1 (1) AUC=0 p=1	no_data

- Improved report Download as excel with additional formatting to clarify predictions on target substance from actual data for source analogues.
- Report downloadable as xlsx file format



- Predictions for target substance are updated in the Data Matrix view upon clicking "Run Read-Across"
- Pagination to facilitate easier scrolling through study effects

Generalised Read-Across (GenRA) Prediction: Updating the Dashboard implementation and using genra-py as a standalone application

Abstract 232

I Shah[1], T Tate[1], L Girardin[2], S Hamilton[2], B Sharma[1], TN Brown[3], K Baron Furuyama[1], P Do[1], G Patlewicz[1*]
[1] Center for Computational Toxicology & Exposure (CCTE), U.S. Environmental Protection Agency, RTP, NC; [2]GDIT, Falls Church, VA; [3]CCTE, U.S. EPA, Duluth, MN

www.epa.gov

MAIN RESULTS: genra-py

Targeted towards users familiar with using Python

- Facilitates GenRA to be applied on users own data. Datasets need to be imported as flat files.
- Standalone python package written in Python 3.8
- Freely available from https://github.com/i-shah/genra-py under MIT license
- Two main ways of accessing the package – users can clone the repository – Python3, Anaconda and Jupyter are required. Additional configuration is captured in the genra-py conda environment file
- A quick install is available by installing the package through pip – 'pip install genra'
- A Docker image is also available at https://hub.docker.com/r/patlewig/genra-py The image comes preinstalled with the Jupyter Scipy stack and the pip installable version of genra-py
- The genra-py repository includes notebooks showcasing how GenRA was used to replicate two of our published analyses in Shah et al (2016) on hazard screening level predictions using ToxRefDB repeated dose data and acute lethality (LD50) in rodents Helman et al (2019).
- A separate github repository is available with the acute toxicity example notebook but which can be run using Binder. This builds and runs a Docker container allowing users to test out the application and work through the notebook without any set up – see https://github.com/patlewig/UNC_Rax

Generalised Read-Across (GenRA) Prediction: Updating the Dashboard implementation and using genra-py as a standalone application

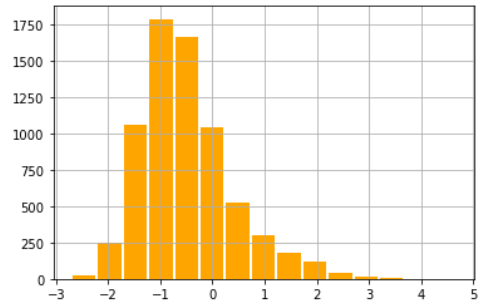
Abstract 232

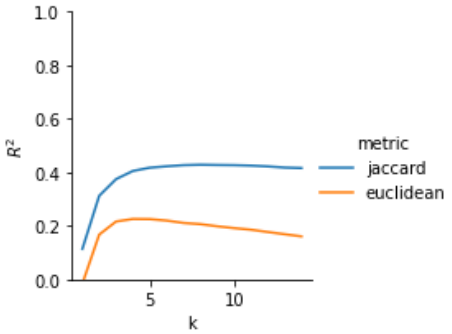
I Shah[1], T Tate[1], L Girardin[2], S Hamilton[2], B Sharma[1], TN Brown[3], K Baron Furuyama[1], P Do[1], G Patlewicz[1*]
[1] Center for Computational Toxicology & Exposure (CCTE), U.S. Environmental Protection Agency, RTP, NC; [2]GDIT, Falls Church, VA; [3]CCTE, U.S. EPA, Duluth, MN

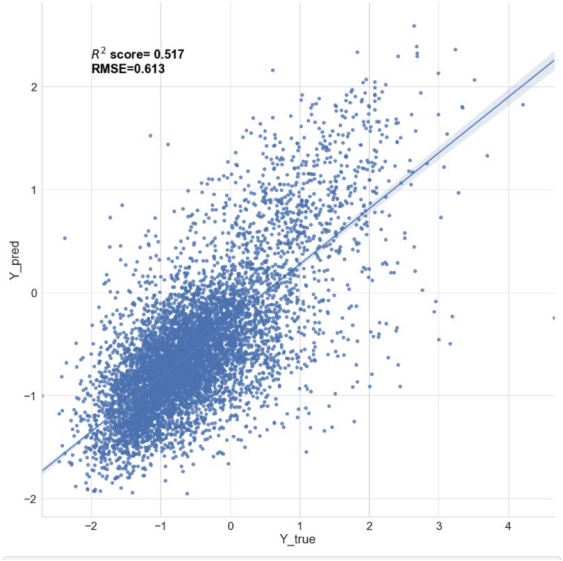
www.epa.gov

MAIN RESULTS: CASE STUDY using genra-py

- Import the acute toxicity data and chemical structure information
- Compute Morgan Fingerprints using the python RDkit library
- View the distribution of the acute lethality data expressed as LD50 in mg/kg but converted to log(molar) equivalents



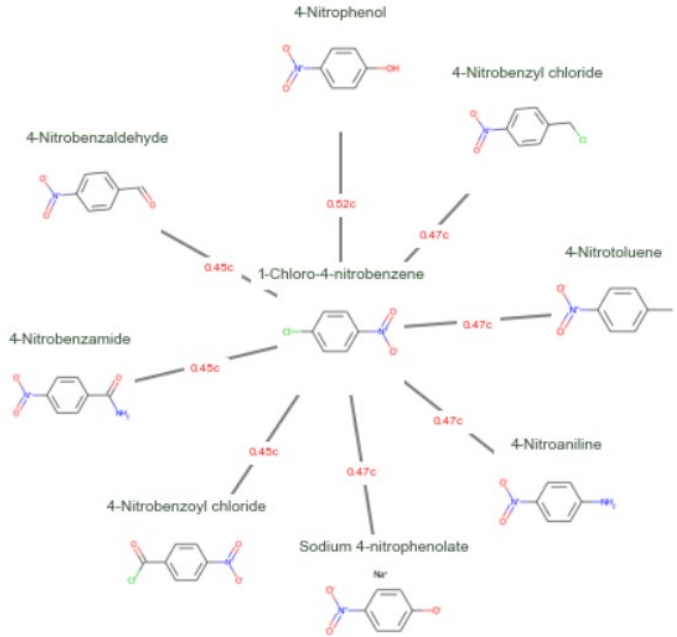




- Calculate the coefficient of determination and root mean squared error to provide a global performance metric for the acute toxicity dataset

- Import genra-py and explore through a grid search what the optimal number of neighbours and similarity metric is for the dataset

- Make a prediction for a specific chemical by identifying the 8 most similar source analogues.



Generalised Read-Across (GenRA) Prediction: Updating the Dashboard implementation and using genra-py as a standalone application

I Shah[1], T Tate[1], L Girardin[2], S Hamilton[2], B Sharma[1], TN Brown[3], K Baron Furuyama[1], P Do[1], G Patlewicz[1*]
[1] Center for Computational Toxicology & Exposure (CCTE), U.S. Environmental Protection Agency, RTP, NC; [2]GDIT, Falls Church, VA; [3]CCTE, U.S. EPA, Duluth, MN

IMPACT/SIGNIFICANCE

- Two implementations of GenRA are now available – an updated web application with the latest underlying chemistry and toxicity data as well as a standalone python package – genra-py.
- We anticipate these versions will facilitate new features and functionalities commensurate with our recent research findings to be more efficiently implemented.
- See posters 231 and 245 by W Jenkins and T Tate which showcase some of our ongoing research in this field.