Influence of Transcriptomic Descriptors on the Generalized Read-Across (GenRA) Performance

T. Tate[1], G. Patlewicz[1], J. Wambaugh[1], I. Shah[1]

[1]Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Abstract 245

Tia Tate | tate.tia@epa.gov | 0000-0002-7359-1660

Objectives

- Read-across is a data gap filling technique utilized to predict the toxicity of a target chemical using data from similar analogues.
- GenRA attempts to make automated read-across predictions of toxicity outcomes for untested chemicals. Previous efforts utilizing GenRA focused on the enhancement of read-across prediction through characterization of chemical structure, physiochemical, and bioactivity similarity.
- In this work, the feasibility and performance of targeted High-Throughput Transcriptomic (HTTr) data in assigning *in vivo* toxicity read-across predictions was investigated.

Approach

- Binary 'hit calls' for targeted HTTr data on 93 transcripts for 1060 chemicals in HepaRG™ cells that measure nuclear receptor activation, xenobiotic metabolism, cell stress, cell cycle progression, and apoptosis, were derived from TCPL multiple concentration level 5 data.
- Newly developed python package, genra-py was utilized to evaluate the baseline chemical structure fingerprint approach versus these transcriptomic fingerprints as well as a hybrid combination of both for predicting ToxRefDB v2.0 hazard outcomes using Area under the Receiver Operating Characteristic (ROC) Curve (AUC).

Key Results

- For all endpoints, there were modest improvements in ROC AUC scores of 0.01 (2.1%) and 0.04 (7.3%) with transcriptomic and hybrid descriptors, as compared to baseline (Morgan) chemical structure fingerprints.
- For liver-specific toxicity endpoints, ROC AUC scores improved by 10% and 17% for transcriptomic and hybrid descriptors, respectively.
- Thus, significant differences appear to be either study type or target specific.

Impact/Significance

- Previously, the GenRA approach was shown to be effective in characterizing source analogues using binary and quantitative measures for chemical and bioactivity descriptors, as well as physiochemical property information.
- In this present study, we found that in specific cases, the utility of transcriptomic hit-call information in addition to chemical structure information was promising in making *in vivo* toxicity predictions.
- While the coverage of this dataset was limited, we anticipate that these and future expansions to more comprehensive transcriptomic datasets will provide stronger justification for the use of NAMs to enhance read across predictions.

U.S. Environmental Protection Agency
Office of Research and Development

Influence of Transcriptomic Descriptors on the Generalized Read-Across (GenRA) Performance

T. Tate[1], G. Patlewicz[1], J. Wambaugh[1], I. Shah[1]

[1]Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Abstract 245

Tia Tate | tate.tia@epa.gov | 0000-0002-7359-1660

OBJECTIVES

❖Given both national and international efforts to significantly reduce animal testing through developing new approach methodologies (NAMs) to inform chemical hazards and risks, our objective was to investigate the feasibility and performance of targeted high-throughput transcriptomics (HTTr) in assigning *in vivo* toxicity read-across predictions for untested chemicals using the Generalized Read Across (GenRA) approach (Shah et al., 2016).

❖In subsequent analyses, we have focused on exploring enhancements to read-across; either through characterizing other similarity considerations, e.g., physicochemical properties as a surrogate for bioavailability (Helman et al., 2018) and quantifying their relative contribution to improving read-across performance or transitioning to predictions of potency (Helman et al., 2019a, b).

❖Our current works utilizes the newly developed *genra-py* package to evaluate whether HTTr descriptors individually or in combination with chemical structure descriptors offer improvement and/or significant benefit in predicting *in vivo* toxicity outcomes.

Influence of Transcriptomic Descriptors on the Generalized Read-Across (GenRA) Performance

T. Tate[1], G. Patlewicz[1], J. Wambaugh[1], I. Shah[1]

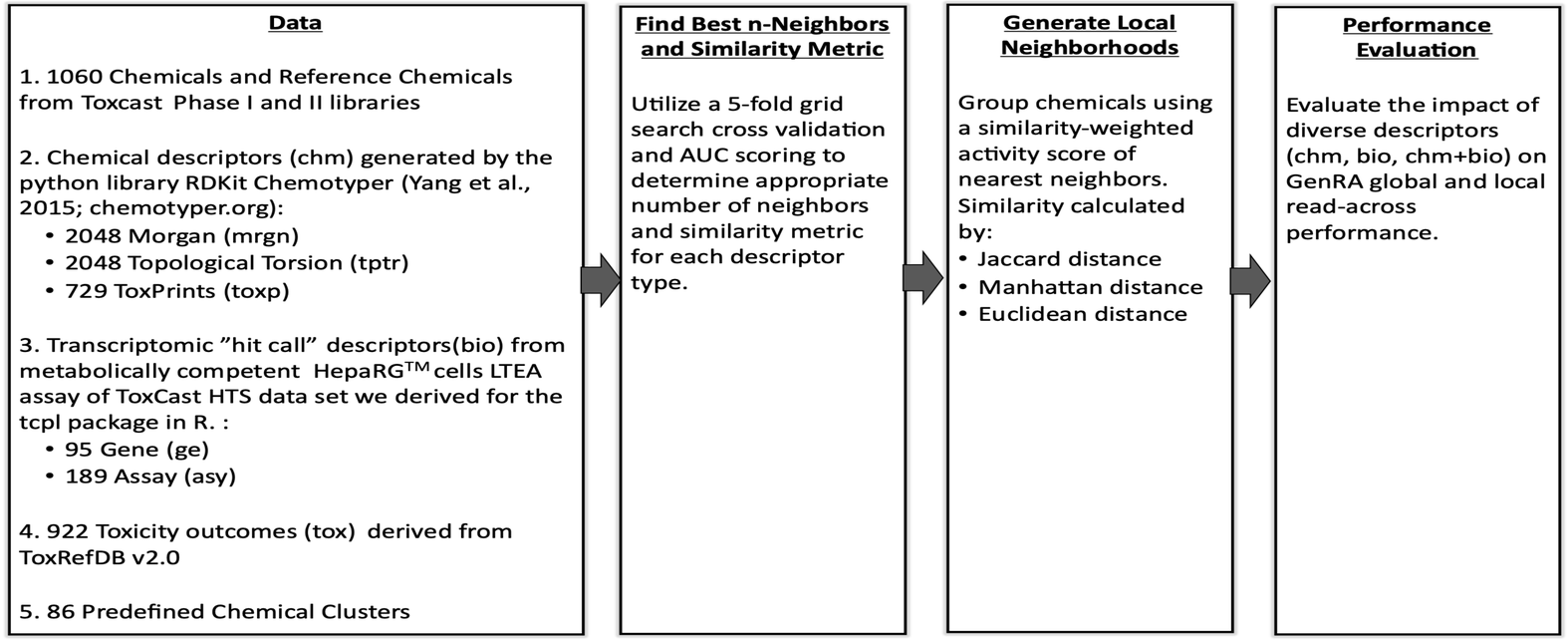
[1]Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Abstract 245

Tia Tate | tate.tia@epa.gov | 0000-0002-7359-1660

Approach

- Binary 'hit calls' for **targeted HTTr data on 93 transcripts for 1060 chemicals in HepaRG™ cells** that measure nuclear receptor activation, xenobiotic metabolism, cell stress, cell cycle progression, and apoptosis, were derived from **TCPL multiple concentration level 5 data.**
- Newly developed python package, **genra-py** was utilized to evaluate the baseline chemical structure fingerprint approach versus these transcriptomic fingerprints as well as a hybrid combination of both for predicting **ToxRefDB v2.0 hazard outcomes using Area under the Receiver Operating Characteristic (ROC) Curve (AUC).**
- The classification accuracy of GenRA for each toxicity endpoint depends on the type of descriptor (chm, bio, and CB descriptors), the choice of similarity metric, and the number of nearest neighbors.



Influence of Transcriptomic Descriptors on the Generalized Read-Across (GenRA) Performance

T. Tate[1], G. Patlewicz[1], J. Wambaugh[1], I. Shah[1]

[1]Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Abstract 245

Tia Tate | tate.tia@epa.gov | 0000-0002-7359-1660

Key Results

Descriptor Type	Descriptor Name	AUC	Metric	Number of Neighbors
Chm	tptr	0.6303	Euclidean	9
Chm	mrgn	0.64549	Jaccard	8
Chm	toxp	0.61379	Jaccard	7
Bio	ge	0.648847	Euclidean	14
Bio	asy	0.6632	Euclidean	11
Hybrid	mrgn + asy (ma)	0.6883	Jaccard	13
Hybrid	toxp + ge (txg)	0.7044	Jaccard	10
Hybrid	tptr + ge (ttg)	0.6818	Euclidean	6
Hybrid	(CB) all	0.6999	Jaccard	14
Chm	(CA) all	0.6702	Jaccard	10
Hybrid	**mrgn + ge (mg)**	**0.7049**	**Jaccard**	**10**
Hybrid	toxp + asy (txa)	0.6992	Jaccard	14
Hybrid	tptr + asy (tta)	0.6721	Manhattan	5

- The classification accuracy of GenRA for each toxicity endpoint depends on the type of descriptor (chm, bio, and Hybrid Chemical and Biological (CB) descriptors), the choice of similarity metric, and the number of nearest neighbors.
- We used genra-py to systematically explore the relationship between these parameters for every toxicity endpoint.
- There was no consensus on the optimum number of neighbors across all descriptor types.
- GenRA parameters used for all toxicity classes was Jaccard similarity metric with 10 nearest neighbors.

Influence of Transcriptomic Descriptors on the Generalized Read-Across (GenRA) Performance

T. Tate[1], G. Patlewicz[1], J. Wambaugh[1], I. Shah[1]

[1]Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Tia Tate | tate.tia@epa.gov | 0000-0002-7359-1660

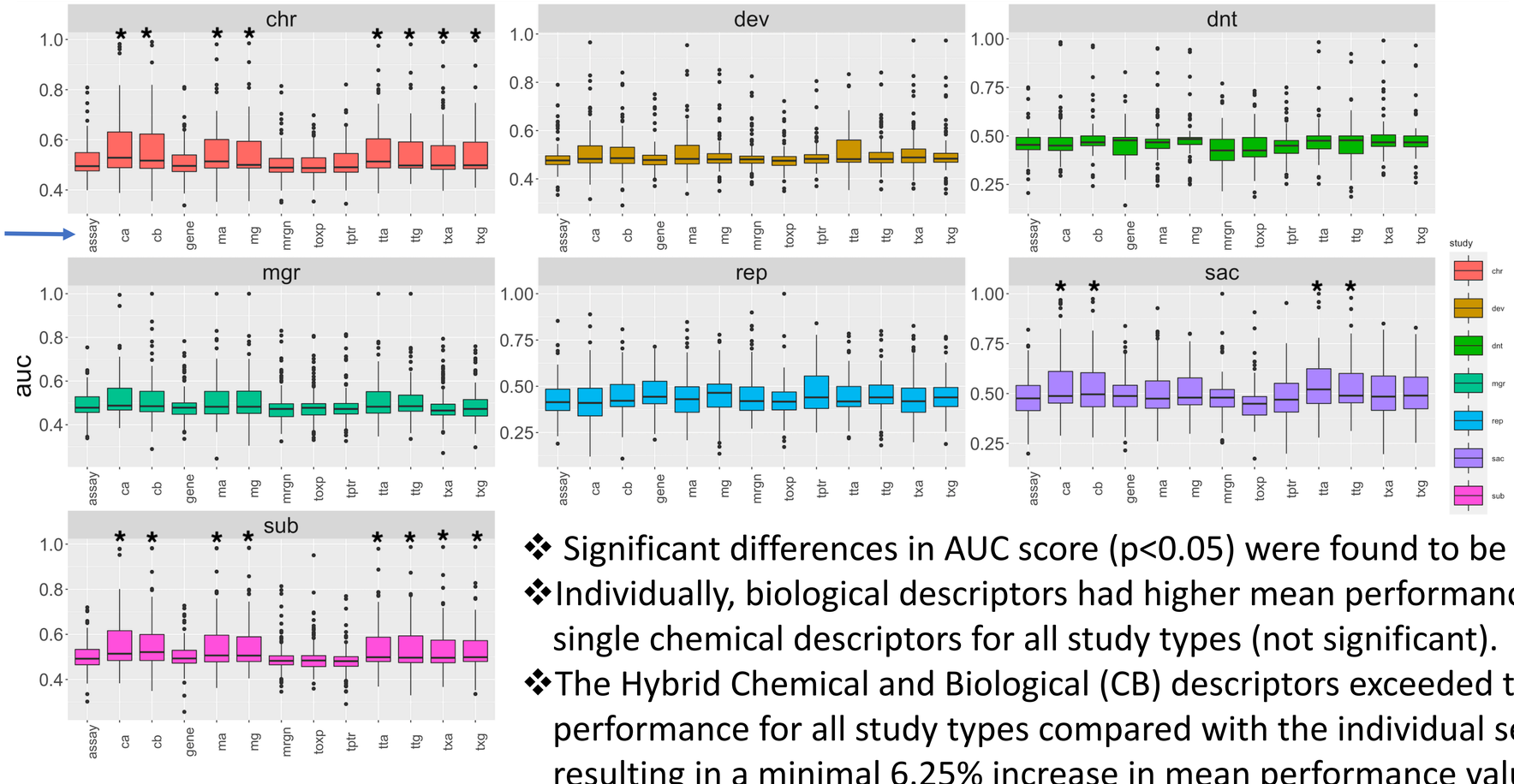
Abstract 245

www.epa.gov

Descriptor Key

Descriptor Type	Descriptor Name
Chm	Torsion (tptr)
Chm	Morgan (mrgn)
Chm	ToxPrints (toxp)
Bio	Gene (ge)
Bio	Assay (asy)
Hybrid	mrgn + asy (ma)
Hybrid	toxp + ge (txg)
Hybrid	tptr + ge (ttg)
Hybrid	All Chm + BIO (CB)
Chm	All Chm (CA)
Hybrid	**mrgn + ge (mg)**
Hybrid	toxp + asy (txa)
Hybrid	tptr + asy (tta)

Key Results



- Significant differences in AUC score ($p<0.05$) were found to be study type specific
- Individually, biological descriptors had higher mean performance scores than the single chemical descriptors for all study types (not significant).
- The Hybrid Chemical and Biological (CB) descriptors exceeded the mean prediction performance for all study types compared with the individual sets of descriptors, resulting in a minimal 6.25% increase in mean performance values overall.

Influence of Transcriptomic Descriptors on the Generalized Read-Across (GenRA) Performance

T. Tate[1], G. Patlewicz[1], J. Wambaugh[1], I. Shah[1]

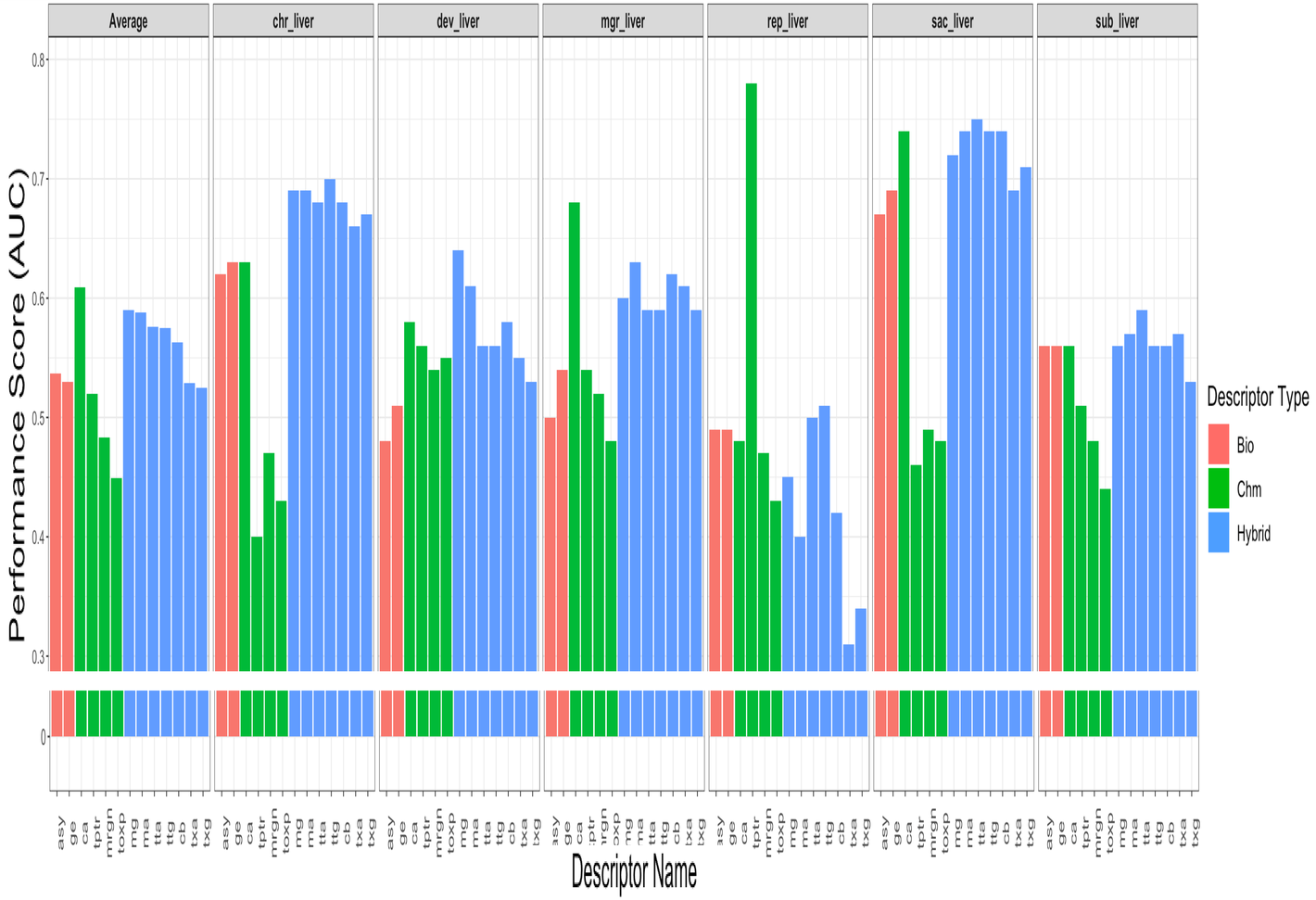
[1]Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Tia Tate | tate.tia@epa.gov | 0000-0002-7359-1660

Abstract 245

Key Results

- For liver-specific toxicity endpoints, ROC AUC scores improved by 10% and 17% for transcriptomic and hybrid descriptors, respectively.
- For chronic, developmental, sub-acute, and sub-chronic liver outcomes, hybrid descriptors resulted in the highest respective significant predictive performance scores.
- The Chemical Combination (CC) descriptors resulted in the best prediction performance (AUC=0.68, $p<0.05$) for multigenerational reproductive liver endpoints.
- Despite the low average AUC scores, biological descriptors produced a modest 10% improvement over chemical descriptors.



Influence of Transcriptomic Descriptors on the Generalized Read-Across (GenRA) Performance

T. Tate[1], G. Patlewicz[1], J. Wambaugh[1], I. Shah[1]

[1]Center for Computational Toxicology and Exposure, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Abstract 245

Tia Tate I tate.tia@epa.gov I 0000-0002-7359-1660

Impact/Significance

- Previously, the GenRA approach was shown to be effective in characterizing source analogues using binary and quantitative measures for chemical and bioactivity descriptors, as well as physiochemical property information.

- In this present study, we found that in specific cases, the utility of transcriptomic hit-call information in addition to chemical structure information was promising in making *in vivo* toxicity predictions.

- While the coverage of this dataset was limited, we anticipate that these and future expansions to more comprehensive transcriptomic datasets will provide stronger justification for the use of NAMs to enhance read across predictions.

Next Steps:

- Evaluation of HTTr in multiple cell types (beyond the liver) for screening thousands of chemicals
- Expansion of scope of biological fingerprints to biological pathways
- Comparison of the GenRA approach to other machine learning approaches