Comparing and combining *in silico* outcomes and *in vitro* mechanism-based assays to predict genotoxicity

Susanne A. Stalford[1], Alex N. Cayley[1], Remco Derr[2] and Giel Hendriks[2].

[1]) Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS, United Kingdom
[2]) Toxys, Robert Boyleweg 4, 2333 CG, Leiden, The Netherlands

Introduction

Genotoxicity is of major concern in human health risk assessment, and as such, regulatory guidance across a diverse number of industries calls for extensive genotoxicity testing to be carried out on new chemical entities. As there is an ambition to abolish *in vivo* experiments when assessing safety, *in vitro* assays have also been incorporated into regulatory decision making.

(Q)SAR and high-throughput assays may be considered as alternatives to such assays. These experiments can give valuable information about mechanism of action, which may also make them useful as screening tools in a discovery context. However, it is not clear how these technologies compare in terms of activity prediction, and if results from these can be combined to give better predictions for genotoxicity? Therefore, in this work, we investigated examples of two such systems to answer these questions.

Methods

Data for 93 compounds, including ECVAM-rated compounds was collated from various experimental systems:

- Derek Nexus [1], well-established piece of software which can give predictions for a variety of toxicity endpoints based on expert-derived structure activity rules. For this analysis, we assessed the compounds against the mutagenicity, chromosome damage and non-specific genotoxicity endpoints using bacterium and mammal species, with the results covering *in vitro* and *in vivo* predictions.
- ToxTracker [2], a unique flow cytometry-based assay that uses biomarkers to identify genotoxic, indirectly genotoxic or non-genotoxic modes of action, including DNA damage, oxidative stress, protein misfolding and general cellular stress mechanisms that are associated with increased cancer risk (Table 1).
- Standard *in vitro* and *in vivo* assays, which in this analysis were combined to give an overall genotoxicity call when a) treating an active outcome in any assay as a positive overall call, or b) using an activity call from *in vitro* data only when there is no *in vivo* data available as a type of "tiered reasoning" approach to validation.
    - *In vitro* regulatory assays (collated from various sources, including Vitic, EPA ACToR and a Kirkland publication [3,4,5]) – Ames, mouse lymphoma assay (MLA), chromosomal aberrations (CA) and micronucleus (MN)
    - *In vivo* regulatory assays (collated from various sources, including Vitic, EPA ACToR and a Kirkland publication [3,4,5]) – CA and MN

The results from Derek Nexus and ToxTracker were compared to assess the similarity in genotoxicity prediction. This included evaluating each of the individual biomarkers and relevant combinations thereof, (e.g. combining results from the Bscl2 and Rtkn reporters to give a DNA damage call or the Blvrb and Srxn1 reporters for oxidative stress, on the basis that a positive outcome in one biomarker constitutes a positive outcome overall) against whether a Derek Nexus genotoxicity alert was activated or not.

The *in silico* and high-throughput results were then analysed against the *in vitro* and *in vivo* data to look at how prediction of genotoxic activity changes when using more than one system (Figure 1).

References

1. https://www.lhasalimited.org/products/derek-nexus.htm
2. Hendriks et al, Toxicol. Sci, 2016, 150,190–203. doi:10.1093/toxsci/kfv323.
3. https://www.lhasalimited.org/products/vitic.htm
4. https://actor.epa.gov
5. Kirkland et al, Mutat. Res., 2016, 795, 7-30. doi: 10.1016/j.mrgentox.2015.10.006

Results

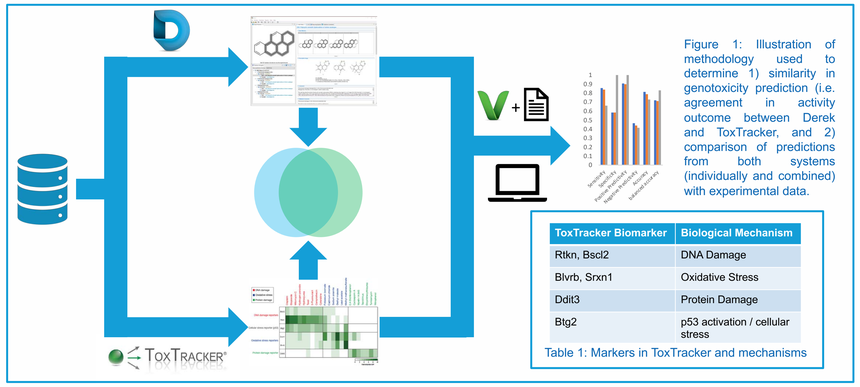

Figure 1: Illustration of methodology used to determine 1) similarity in genotoxicity prediction (i.e. agreement in activity outcome between Derek and ToxTracker, and 2) comparison of predictions from both systems (individually and combined) with experimental data.

ToxTracker Biomarker	Biological Mechanism
Rtkn, Bscl2	DNA Damage
Blvrb, Srxn1	Oxidative Stress
Ddit3	Protein Damage
Btg2	p53 activation / cellular stress

Table 1: Markers in ToxTracker and mechanisms

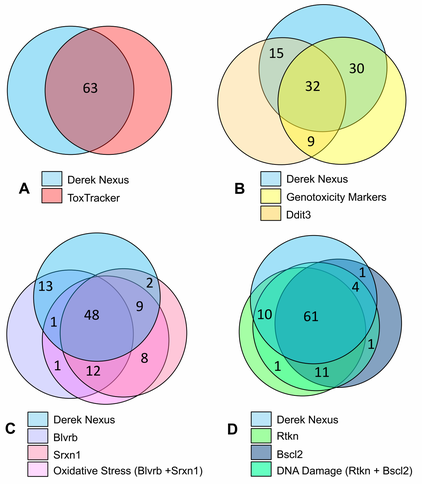

Figure 2: Similarity in predictions between Derek Nexus and ToxTracker. Each circle represents 93 compounds and each number indicates the compounds whose predictions for each system agree with each other. For each combination of biomarkers, a conservative overall call is made (i.e. if positive in one biomarker, an overall call of positive is made).

Comparison of ToxTracker and Derek Nexus showed results from each system agreed ~70% of the time (Figure 2, A). Agreement between results from oxidative stress, DNA damage and cellular stress markers and Derek Nexus was approximately the same as for ToxTracker overall, whereas the Ddit3 biomarker (which does not predict a genotoxic mechanism) results only agreed with Derek ~50% of the time (B). While oxidative stress biomarker results show agreement with Derek Nexus on par with ToxTracker overall (C), Derek Nexus and DNA damage markers give the most similar results, agreeing ~80% of the time (D). This is not unexpected given Derek Nexus genotoxicity alerts are based on data from assays such as Ames, CA and MN which predominantly detect DNA damage. These observations indicate that while both systems can give similar results, it is possible that the two may be used together in a complementary fashion to better predict genotoxicity. However, it is also possible that one system may always be correct, and thus the systems would not be complementary. Therefore the results need to be combined and compared to experimental data to ascertain which hypothesis is correct.

Conclusions and Future Work

These results suggest that combining results from *in silico* and high-throughput biomarker assays can be used to predict *in vivo* genotoxicity, particularly when considering the biomarkers associated with genotoxicity. This combination of systems can thus act as a screening tool to reduce animal testing in a discovery scenario, as well as potentially having applications in other areas of industry. Given the additional mechanistic information and knowledge these systems provide, in the future it may be possible to reason between the results in a more sophisticated way, for example, using an adverse outcome pathway framework, to reach a conclusion which would help combine evidence to waive testing in a regulatory scenario, saving time, money and animals.

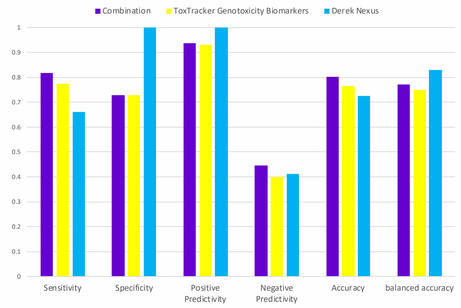

Figure 3: Chart illustrating validation metrics for ToxTracker genotoxicity biomarkers, Derek Nexus and a combination of the results from the two systems.

The results from each system were validated against experimental data from *in vitro* and *in vivo* assays. Initial validation suggests that using a conservative combination of Derek Nexus and ToxTracker oxidative stress, DNA damage and cellular stress markers (where an active outcome in one biomarker or *in silico* result equates to an active outcome overall) predicts better in terms of sensitivity, negative predictivity and accuracy than either system individually. However, Derek Nexus performs better for specificity, positive predictivity and balanced accuracy (Figure 3). This indicates that the two systems can be used in a complementary fashion.

Next, the validation results for the two systems combined was investigated by looking at how the validation results for the two systems in combination varied when only considering individual biomarkers and different combinations of biomarkers from ToxTracker as part of the analysis (again with a positive result in one system or biomarker equating to an active outcome overall). The best validation results are observed when only considering biomarkers associated with DNA damage (Rtkn and Bscl2) and cellular stress (Btg2). While sensitivity drops by 10-14%, specificity greatly increases (~35%) and balanced accuracy increases without a significant loss in accuracy or negative predictivity (Figure 4).

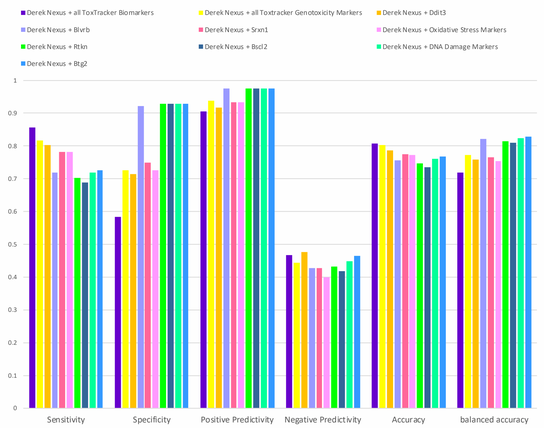

Figure 4: Chart illustrating validation metrics for Derek Nexus in combination with the biomarkers tested in ToxTracker.

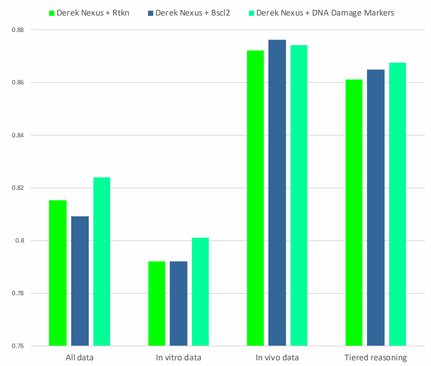

Figure 5: Chart illustrating changes in balanced accuracy for Derek Nexus in combination with DNA damage biomarkers when validating against different combinations of experimental data.

Using only the DNA damage markers and Derek Nexus combined results, the combined results were validated 1) using the overall call of experimental data when *in vitro* and *in vivo* data are treated equally (a positive in one assay equates to positive overall), 2) against *in vitro* and *in vivo* data separately, and 3) by combining the data in a tiered approach (using *in vitro* data only when *in vivo* data is unavailable to determine activity) (Figure 5). Given that the ideal scenario is to predict the *in vivo* genotoxicity of compounds, balanced accuracy was found to be at its highest when using *in vivo* experimental data only. However, as there tends to be less *in vivo* data available, using a tiered reasoning approach (where *in vitro* data is additionally considered only when *in vivo* data is not available) is not significantly detrimental to balanced accuracy. Similar trends are also observed for sensitivity and negative predictivity and accuracy.