Mechanistic Expert Call Datasets Support *in silico* Prediction of Teratogenicity for a Wider Chemical Space

Adrian Fowkes, Jeffery Plante, Alun Myden, Emma Hill, Alex Cayley and M. Bashir Surfraz

Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS

Abstract number - 2619

Exploiting Alternative Data

The scarcity of teratogenicity data and the cost of *in vivo* reproductive toxicity studies are driving the use of a wider range of assays, where the relationship between data and teratogenicity can be established. Similarly this lack of data is affecting the applicability domain of prediction systems for teratogenicity. Using an adverse outcome pathway (AOP) framework, key events (KE) leading to teratogenicity can be mapped and suitable *in vitro* and *in vivo* assays, which model the KEs can be identified. This type of relevant data is available for a significantly larger number of chemicals in comparison to teratogenicity data, which in turn can be mined to extract useful knowledge allowing for teratogenicity predictions for a wider chemical space. This mechanistic approach provides a clear rationale between a specific molecular initiating event (MIE) or KE and a toxicity endpoint.

During this study three AOPs (estrogen receptor modulation (ERM), androgen receptor modulation (ARM) & 5alpha-reductase inhibition (5aRI)) which have a strong association with teratogenicity were mapped. Relevant data for the MIEs and KEs identified were gathered from ChEMBL [1] and curated into structured Lhasa mechanistic expert call activity datasets (LMEADs) (**Fig** 1). These purposeful datasets were then mined for the creation of MIE structural alerts for Derek Nexus, a transparent *in silico* expert system for toxicity prediction [2].

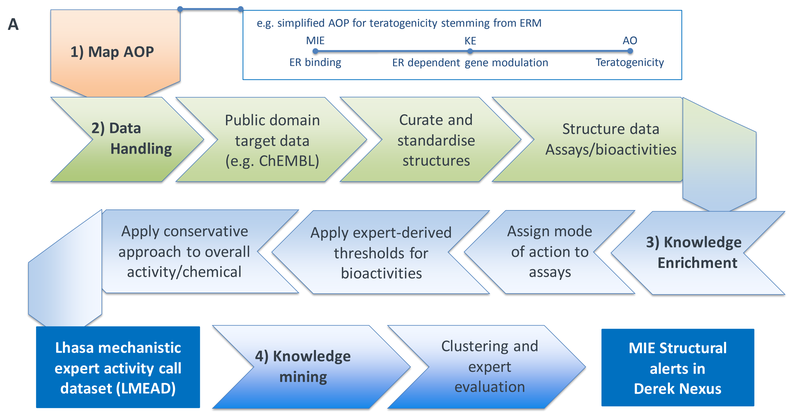

Figure 1A: Workflow describing data handling (in green) followed by knowledge injection (in blue) to create purposeful AOP-themed mechanistic datasets which are then mined for structural alert preparation; **1B**: Illustration of the invaluable input from experts to create a Lhasa ERM dataset from the heterogeneous data downloaded from ChEMBL.

Application in a Transparent Expert System

Mechanistic expert datasets were created for three MIEs (ERM, ARM and 5aRI), all of which were enriched with the injection of additional expert knowledge (**Table 1**). Using the ERM dataset for illustration, activity (active or inactive) was assigned to 85% of compounds in the dataset. In terms of mode of action (agonism or antagonism), 51% of active chemicals were assigned a response type, for which 16% were only attributed a response type upon expert investigation. Furthermore, following expert review 95% of equivocal data points examined were assigned an activity call.

Table 1: Analysis of the LMEADs generated from a semi-automated methodology and significant value-adding from experts to improve quality and accuracy.

LMEAD	Number of substances	Active substances (Equivocals removed)	Response type known for active substances	Data points verified
ERM	6952	53% (55%)	51%	46%
ARM	4849	57% (62%)	68%	64%
5aRI	1261	66% (83%)	NA	66%

For alert implementation the LMEADs were processed against the relevant *in vivo* teratogenicity alerts in Derek Nexus (2014 certified knowledge base) to identify false negatives (FNs - active chemicals in the dataset which failed to activate the relevant *in vivo* teratogenicity alerts). New MIE structural alerts were implemented into Derek Nexus and also incorporated into a custom knowledge base (KB). Potential MIE alerts, in the form of expert-weighted structural hypothesis (EWSHs) were also implemented into each custom KB (**Table 2**). Overall 62%, 58% and 78% of FNs from the ERM, ARM and 5aRI LMEADs respectively were accounted for by each custom KB.

Table 2: Three Derek Nexus custom knowledge bases providing teratogenicity predictions for a considerably wider chemical space based on work completed on relevant MIEs.

Endpoint	Alerts in Custom KB	Number of existing teratogenicity alerts	Number of new MIE alerts	Number of new EWSH	FNs correctly predicted
ERM	48	3	9	36	62%
ARM	23	2	7	14	58%
5aRI	24	1	16	7	78%

MIE-based endpoints are now present in the 2015 certified KB of Derek Nexus for 5aRI (16 alerts) and ERM (9 alerts). Much effort has gone into making their association with teratogenicity clear with the use of extrapolated reasoning (**Fig 2**). Activation of any MIE alert automatically generates a prediction for teratogenicity at an equivocal level, based on a specific reasoning rule implemented for this purpose while providing a clear rationale to justify this association.

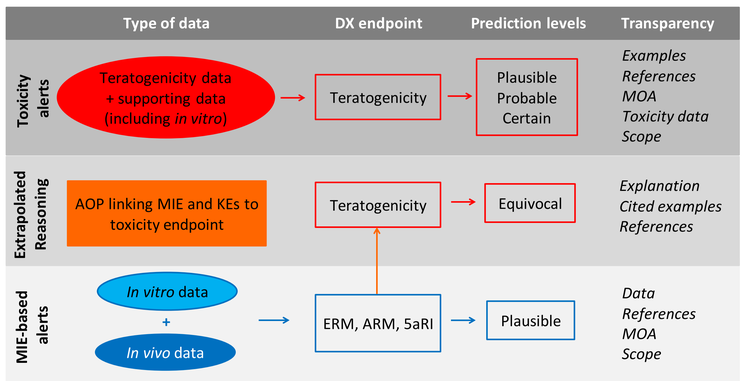

Figure 2: Association between MIE-based endpoints and teratogenicity in Derek Nexus: predicting for a wider chemical space based on alternative data is balanced with a consistent level of transparency in the prediction executed via reasoning rule description comments.

Performance and Validation

The LMEADs for ERM, ARM and 5aRI were processed against the relevant *in vivo* teratogenicity alerts in Derek Nexus (2014 certified KB, referred to as KB1) and the new individual custom KBs (referred to as KB2) (**Fig 3A**). The new MIE alerts and EWSHs resulted in increased coverage of the active compounds in the datasets.

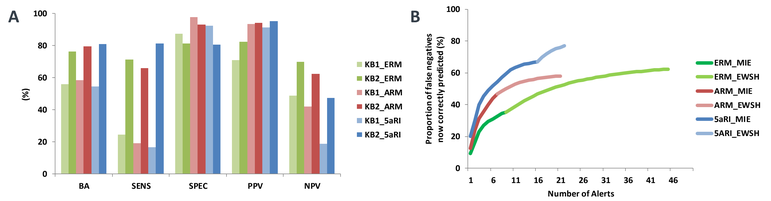

Figure 3A: Performance of each custom KB (KB2) created for the 3 MIEs compared to the relevant Derek Nexus teratogenicity alerts (KB1 - 2014 certified KB); **3B**: Coverage of FNs initially identified in each LMEAD following the implementation of MIE alerts and EWSHs.

Validation has been performed with the ERM custom KB using three different data sources (FDA EDKB [3], ToxCast [4], Tox21 [5]). The datasets were curated by applying the same methodology as described previously. Analysis of the proportion of actives in the dataset showed that the datasets were skewed considerably towards inactive compounds (**Fig 4A**). The custom KB performed well against the FDA EDKB dataset and it was quite encouraging that the expert system scored highly in terms of negative prediction (specificity and negative predictivity), considering the LMEAD is biased towards active compounds. In addition, an evaluation set was created from all three data sources by combining compounds which were absent from the training set (LMEAD). The ERM evaluation set was processed against the ERM custom KB (**Fig 4B).** Considering the fact that the evaluation set is heavily skewed by inactive compounds the relatively low sensitivity score was to be expected. Nonetheless the expert system performed well overall, especially in terms of balanced accuracy, specificity and negative predictivity. Similar validation exercises for the ARM and 5aRI custom KBs are currently in progress. The next step is to assess each custom KB against proprietary datasets from Lhasa Limited members.

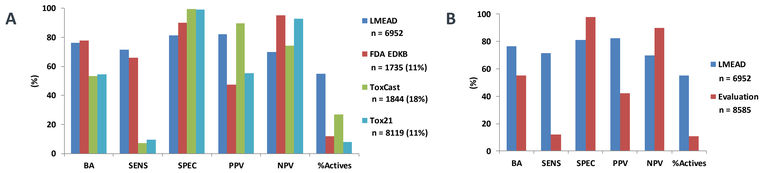

Figure 4A: Performance of the ERM custom KB with datasets derived from different sources; **4B**: Performance of the ERM custom KB with the ERM training set (LMEAD) and the ERM evaluation set. n = number of substances in the dataset with the proportion of substances present in the LMEAD in parenthesis.

Prospective

The Lhasa mechanistic datasets, with expert activity calls assigned to each chemical, are useful both for the development of expert transparent systems and machine learnt models. Furthermore, these MIE-specific datasets can be housed in a purposeful database which can support mining of AOP-derived knowledge, where toxicity can be associated with the relevant assay data.

Future work will focus on glucocorticoid receptor modulation and potentially aromatase inhibition as additional MIE endpoints for the knowledge-based toxicity prediction system – Derek Nexus. This will facilitate an improvement in terms of coverage for transparent teratogenicity predictions, which will be based on both *in vivo* teratogenicity data and relevant pharmacological data.

References

[1] Bento *et al*, Nucleic Acids Res., 42, 2014, 1083-1090, [2] Derek Nexus - Lhasa Limited, [3] Ding *et al*, BMC Bioinformatics, 11, 2010, S5, [4] EPA - ToxCast, [5] NIH – Tox21