Derek Nexus predicts human skin sensitisation accurately. What is the rationale behind its predictive performance?

Sebastien Guesne, Steven Canipa and Richard Williams.

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

I - Introduction

Skin sensitisation is a toxicological endpoint and results in allergic contact dermatitis (ACD) which is the clinical manifestation of an allergy to a topically applied substance.[1] As a consequence of the social and economical cost of ACD, the evaluation of the sensitisation potential of substances is required by a variety of regulators world-wide. Although this assessment has traditionally been conducted using animal assays – such as the Guinea Pig Maximisation Test (GPMT)[2] and the Local Lymph Node Assay (LLNA)[3] – concerted efforts have recently been made to develop in vitro alternatives.

Examples of tests currently under formal validation at the European Union Reference Laboratory for alternatives to animal testing (EURL ECVAM) include the direct peptide reactivity assay (DPRA)[4], the KeratinoSens assay[5] and the human cell line activation test (h-CLAT).[6]

The validation of these new in vitro tests often involves comparing results against data taken from animal studies, typically the LLNA or GPMT assays. However, when considering the suitability of new tests for skin sensitisation assessment it is important to also consider performance against human data since it is the potential for toxicity in humans that is ultimately of interest.

With this in mind, we decided to investigate how accurately Derek Nexus[7] – an SAR-based expert-system for the prediction of toxicity – predicted a dataset of substances with well identified hazards for the endpoint of human skin sensitisation.

II - Method

Human skin sensitisation data were retrieved from 3 public sources, resulting in three individual datasets:

i) - The Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV) list - Federal Institute for Risk Assessment since November 2002 (BfR).[8] The resulting **BgVV dataset** contained 255 substances in which 183 and 72 were classified as positive and negative respectively.
ii) - A publication by Basketter et al. in which 131 substances were classified in 6 categories based on literature analysis by the authors.[9] The resulting **Basketter dataset** provided 124 substances in which 76 and 48 were classified as positive and negative respectively.
iii) - A search of the eChemPortal database for compounds submitted to ECHA under REACH.[10] The resulting **eChemPortal dataset** provided 98 substances in which 15 and 83 were classified as positive and negative respectively.

The three data sets were then combined. Ten substances had conflicting data and were discarded to yield a final **combined human dataset** of 404 unique substances with 222 positive and 182 negative results.

LLNA[11] and GPMT[12] data were retrieved from the literature and Vitic Nexus database[13] and curated at Lhasa Limited. The final LLNA and GPMT datasets consisted of 398 and 198 substances respectively which were classified as sensitisers (positive) or non-sensitisers (negative). LLNA and GPMT datasets had respectively **91** and **53** substances in common with the combined human dataset - 63/28 and 36/17 positive/negative respectively. Compounds were processed through Derek Nexus.[7] Derek Nexus predictions of at least Equivocal were classified as positive, other predictions and "nothing to report" were classified as negative.

III - Results

1 - Derek Nexus predictive performance against individual datasets

Figure 1 shows contingency tables of Derek Nexus predictions against the 3 individual datasets and Chart 1 the corresponding Cooper statistics. The performance of Derek Nexus varied depending on the dataset used. While high sensitivity but low specificity was observed for the BgVV dataset, the opposite was true for the eChemPortal dataset. Derek Nexus displayed good balanced accuracy (66-76%) against all 3 datasets.

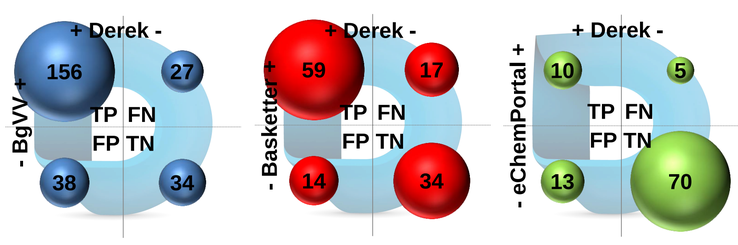


Figure 1. Contingency tables of Derek Nexus predictions against the individual datasets

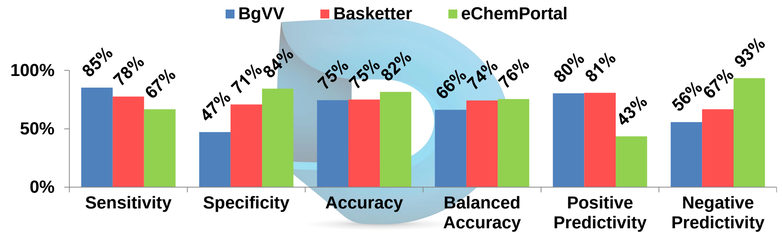


Chart 1. Derek Nexus performance figures against BgVV, Basketter and eChemPortal datasets

2 - Derek Nexus predictive performance against the Combined Human Dataset

Figure 2 shows the contingency table of Derek Nexus predictions against the combined human dataset of 404 unique substances and Graph 1 the corresponding Cooper statistics.

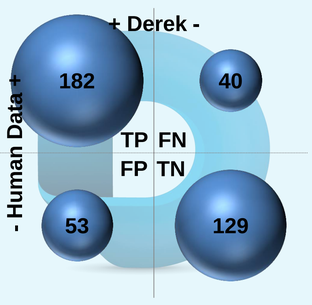
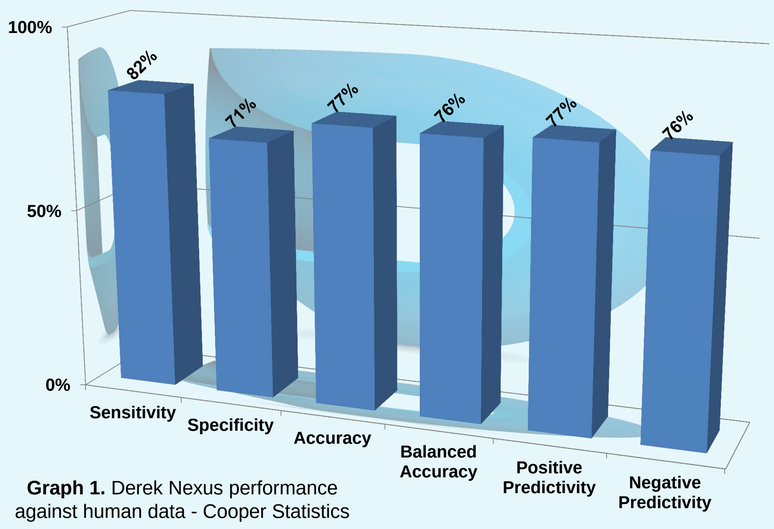


Figure 2. Contingency table of Derek Nexus predictions against the combined human dataset



Graph 1. Derek Nexus performance against human data - Cooper Statistics

Derek Nexus predictive performance against the combined human dataset was good: scores of 82% for sensitivity and 71% for specificity were found. In addition accuracy, balanced accuracy, positive predictivity and negative predictivity were all above 75%.
Derek Nexus contains a total of 73 alerts for the endpoint of skin sensitisation. With the combined human dataset, 47 of these alerts were activated (64% of all the skin sensitisation alerts) with an average positive predictivity of 79%. This illustrates that the human dataset covered a wide range of the Derek Nexus knowledge base (Figure 3).

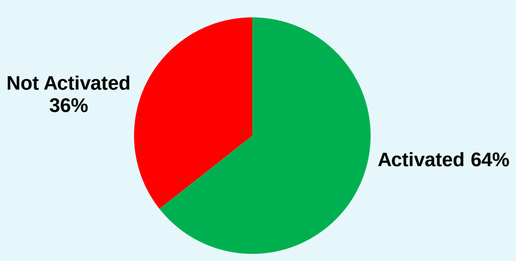
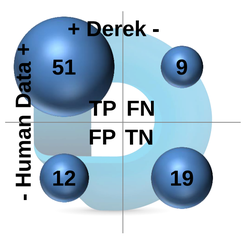


Figure 3. Alerts covered by the combined human dataset

3 – Performance of Derek Nexus and the LLNA against Human Data



Vs

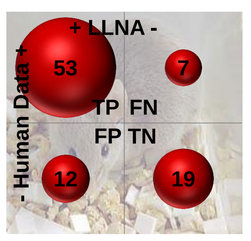
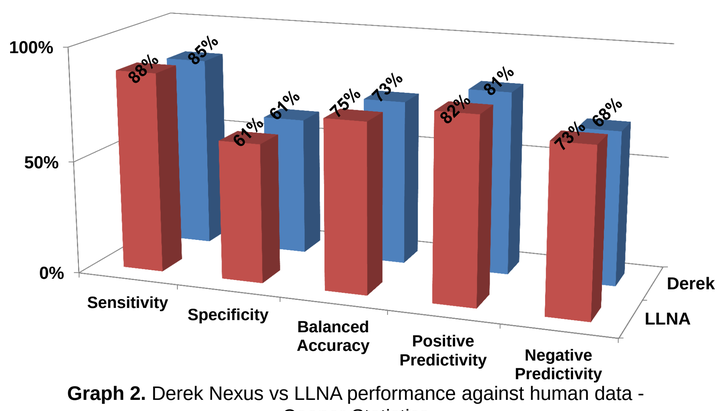


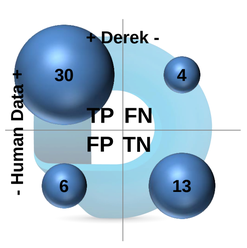
Figure 4. Contingency table of Derek Nexus vs LLNA predictions against human data

When Derek Nexus was compared to the LLNA to predict the skin sensitisation hazard against the human dataset of **91** substances, the following contingency tables (Figure 4) and corresponding Cooper statistics (Graph 2) were found. The predictive performance of Derek Nexus and the LLNA against the human data were very similar with each test displaying high sensitivity but relatively low specificity.



Graph 2. Derek Nexus vs LLNA performance against human data - Cooper Statistics

4 – Performance of Derek Nexus and the GPMT against Human Data



When Derek Nexus was compared to the GPMT to predict the skin sensitisation hazard against the human dataset of **53** substances the following contingency tables (Figure 5) and corresponding Cooper statistics (Graph 3) were found. Once again, high sensitivity but low specificity was observed for both tests although Derek Nexus performed slightly better than the GPMT.

Vs

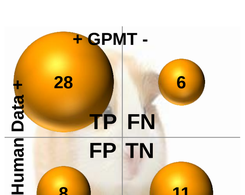
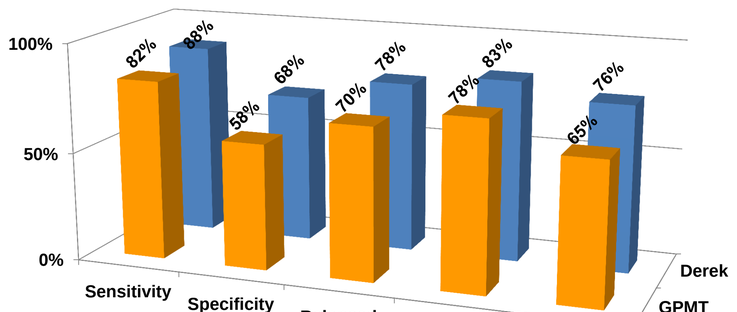


Figure 5. Contingency table of Derek Nexus vs GPMT predictions against human data



Graph 3. Derek Nexus vs GPMT performances against human data - Cooper Statistics

IV - Conclusion

This study shows that Derek Nexus is capable of predicting human skin sensitisation datasets with good balanced accuracy (66-76%). This is supported by the performance figures against the combined human dataset in which a high balanced accuracy of 76% was found.
The performance of Derek Nexus against the combined human dataset was similar to that observed with in vivo assays with each displaying high sensitivity (82-88%) and relatively low specificity (58-71%). One explanation for this may be the fact that Derek Nexus alerts for skin sensitisation have been derived primarily using data from animal assays, such as the LLNA and the GP assays as illustrated in Figure 5. These in vivo assays are often conservative tests and have a tendency to generate false positive predictions when measured against human data. It is therefore reasonable to assume that the alerts in Derek Nexus have inherited the high sensitivity of the animal studies.

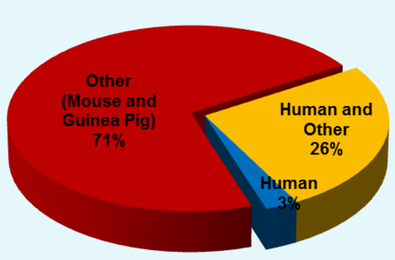


Figure 5. Alerts categorised according to species used in supporting toxicity data

References

[1] Karlberg et al. *Chem. Res. Toxicol.* **28**, 53-69 (2008); [2] *OECD*, Test Guideline No. 406: Skin Sensitisation (1992); [3] *OECD*, Test Guideline No. 429: Skin sensitisation: Local Nymph Node assay (2010); [4] a) Gerberick et al. *Toxicol. Sci.* **81**, 332-343 (2004); b) Gerberick et al. *Toxicol. Sci.* **97**, 417-427 (2007); [5] Natsch et al. *Toxicol. Sci*. **107**, 106-121 (2009); [6] a) Sakaguchi et al. *Toxicol. In Vitro* **20**, 774-784 (2006); b) Ashikaga et al. *Altern. Lab. Anim.* ***38***, 275-284 (2010); [7] Derek Nexus, version 3.0, Lhasa Limited, Leeds, UK, 2012; [8] a) Kayser, D., Schlede, E. (Eds.), 2001. Chemikalien und Kontaktallergie—Eine bewertende Zusammenstellung. Verlag Urban & Vogel, München; ISBN 3-86094-163-1; b) Schlede et al. *Toxicology* **193**, 219-259 (2003); [9] Basketter et al. *Dermatitis* **25**, 11-21 (2014); [10] http://echa.europa.eu/; [11] a) Gerberick et al. *Dermatitis* **16**, 157-202 (2005); b) Kern et al. *Dermatitis* 21, 8-32 (2010); c) Natsch et al. *J. Appl. Toxicol.* **33**, 1337-1352 (2013); d) Nukada et al. *Toxicol. In Vitro* **27**, 609-618 (2013); e) Teubner et al. *Regul. Toxicol. Pharmacol.* ***67***, 468-485 (2013); [12] Cronin and Basketter *Environ. Res.* ***2***, 159-179 (1994); [13] Vitic Nexus, version 2.0.0, Lhasa Limited, Leeds, UK, 2011.