Integrated hazard identification of chemical sensitizers using in vitro and in silico readouts – A comparative evaluation of predictive performance

Donna Macmillan1, Henrik Johansson2, Olivia Lame2, Malin Lindstedt3
1. Gramarye Wharf House, 2 Canal Wharf, Leeds, LS11 5PS
2. SenzaGen, Lund, Sweden
3. Lund University, Lund, Sweden

Abstract no. 322

Introduction

There has been a significant drive to reduce, refine and replace animal models for the prediction of skin sensitization. This is in part due to the implementation of EU regulation 1223/2009 which prohibits the sale and marketing of any cosmetics and cosmetic ingredients which have been tested on animals, alongside REACH® and CLP® regulations which state that non-animal methods must be exhausted prior to considering the use of animal tests. The use and availability of non-animal methods is ever-increasing and 3 assays have been validated by the OECD thus far; the in chemico DPRA, the in vitro KeratinoSens™ and the in vitro h-CLAT. A number of other assays are undergoing OECD validation, including the GARDskin assay (Genomic Allergen Rapid Detection), a dendritic cell-based assay which identifies skin sensitizers from 200 genomic biomarkers4. However, it is generally accepted that no single non-animal method can be used as a standalone approach to replace animal models such as the murine local lymph node assay (LLNA). The focus has instead turned to combining multiple in chemico/in vitro assays and/or molecular descriptors to derive a more accurate assessment of hazard or risk, known as integrated testing strategies (ITS)5. The GARDskin assay has demonstrated high predictivity and has been reported as ready to use in an ITS, therefore, it was decided to investigate the effect on performance when GARD was used in combination with Derek Nexus - and to compare these results against Derek with the DPRA, KeratinoSens™ and h-CLAT.

How does the integration of Derek affect in chemico/in vitro assay performance against the LLNA?

Performance using in chemico/in vitro assays or in silico methods alone - The predictive performance is largely comparable across the four in chemico/in vitro assays being evaluated in this study and Derek - although h-CLAT, GARDskin and Derek demonstrate slightly superior results to DPRA and KeratinoSens™ (Figure 1). Derek correctly predicts non-sensitizers more often than the in chemico/in vitro assays. The number of false negatives i.e. sensitizers incorrectly classified as non-sensitizers, is of particular importance where alternative methods are used in lieu of animal tests for establishing human safety, as the consequences (consumers being exposed to sensitizers) are severe. Accordingly, Derek predictions were used together inside the chemico/in vitro assay result to try and improve negative predictivity and improve overall performance.

Performance using Derek and in chemico/in vitro assays - Derek predictions were incorporated with the assay results using a conservative call approach - if either the Derek prediction or the assay result was positive then the chemical was classified as a sensitizer. If both Derek and the assay were negative then the chemical was classified as a non-sensitizer. Combining a Derek prediction with one assay has a mostly beneficial effect on performance (Figure 2) especially accuracy, sensitivity and negative predictivity, compared to using in chemico/in vitro assays alone (Figure 1). This has been at the expense of specificity which has fallen by 20-25% (Table 1) for most assay combinations (Derek + DPRA, KeratinoSens™, or h-CLAT) - however, an increase in false positives, and concomitant drop in specificity and predictivity is not unexpected from a conservative approach. For this dataset, using Derek with GARDskin demonstrates the best overall results (Figure 2 and Table 1).

Affect on number of mispredictions - Combining Derek and assay results can reduce the number of overall mispredictions (mispredictions = sensitizers classified as non-sensitizers and vice versa) significantly. DPRA mispredictions decrease by 40%, KeratinoSens™, 36%, h-CLAT, 26%, and GARDskin, 46% (Figure 3).

Does GARDskin predict well against human data?

GARDskin (n = 57) displays a similar predictivity performance towards human sensitization data as both Derek (n = 113) and the LLNA (n = 105) (Figure 4). Direct comparison is complicated by differences in data size and composition e.g. the GARDskin dataset has very few non-sensitizers which significantly impacts the specificity. Nevertheless, addition of Derek to GARDskin improves performance by correctly predicting sensitizers benzoyl peroxide, lauryl gallate, thioglycerol and methymethacrylate, all incorrectly predicted as non-sensitizers (Table 2).

Conclusion

Using Derek skin sensitization predictions in combination with in chemico/in vitro assay results has a beneficial effect when predicting the LLNA outcome. GARDskin in particular performs extremely well when used with Derek in a conservative call approach. Human sensitization is more challenging to predict and GARDskin performs well for this compared to predicting the LLNA - attributed to the small number of chemicals with both GARDskin and human data (n = 57), in addition to the positive bias in the GARD dataset (70%). However, the addition of Derek predictions clearly improve assay performance. Future work will focus on repeating this analysis on a larger, more balanced dataset.

References: