



Letter to the editor regarding the article by Roberts, 2018

To the editor

In a recent contribution by D. Roberts, funded by the developers of the GARD™ assay, the question was raised: „Is a combination of assays really needed for non-animal prediction of skin sensitization potential?” (Roberts, 2018) and it was proposed that, based on the superior predictivity on a set of chemicals tested by the assay developers, the transcriptomic GARD™ assay as stand-alone assay may be a better option as the currently widely discussed proposal of using multiple test covering different steps of the Adverse Outcome Pathway (AOP) (OECD, 2012). A strong argument for a stand-alone assay instead of a combination of tests, namely a ‘2 out of 3’ approach, had been made by the GARD™ development team in an earlier review, too (Johansson and Gradin, 2017). This proposal and the development of the GARD™ assay challenges a number of conventions in *in vitro* toxicology, challenges we think should be carefully discussed.

1. Challenging the AOP approach

The author makes the correct observation, that strict ‘AOP logic’ – describing the sensitization phase as a linear process covered by mechanistic steps – would require a series of *in vitro* tests, and all of those would need to be tested positive for a positive final outcome (Roberts, 2018). This linear view is actually the proposal made for the first time in the seminal paper by Jowsey et al. (2006), which predates the AOP discussion but which is very much recapitulated in the way the skin sensitization AOP was drawn by the OECD (OECD, 2012). This approach would always lead to very high specificity, but low sensitivity. However, it is widely acknowledged, that AOPs should not be viewed in such a linear way. Using a suite of mechanistic tests within an AOP may also be viewed as collecting mechanistically relevant data on defined molecular endpoints to finally make a ‘weight of evidence’ judgement. For skin sensitization, it is widely acknowledged that this WoE can be derived from chemical information, reactivity measurements describing the molecular initiating event (MiE) combined with cell based assays.

The assays and *in silico* tools developed to test chemicals for skin sensitization potential are all based on a good, albeit not complete, understanding of the molecular mechanisms underlying skin sensitization. The development of the AOP approach in toxicology in general is based on the assumption that increased mechanistic understanding and testing of defined molecular endpoints will lead to continuously improved prediction of apical endpoints.

The GARD™ assay is only using a single cell-type and (like some related assays based on a pan-transcriptomic analysis) is based on an unsupervised transcriptomic approach not taking any *a priori* mechanistic knowledge into account but defining the measured endpoints

(transcriptomic signature) completely based on statistical evaluation of gene expression changes induced by a set of sensitizers and non-sensitizers. On the positive side, by measuring multiple transcripts not previously assessed, information content is greatly increased and any bias from previous research is removed – however, using this assay as stand-alone assay as proposed would also ignore all previous mechanistic learnings in skin sensitization and statistics would dominate mechanistic knowledge – which of course is fine if indeed our mechanistic understanding is inferior to a purely statistical evaluation of ‘big data’. However, to miss the opportunity to carefully characterize the MiE with *in chemico* methods and trust that chemicals will react with all relevant proteins in the cells and then trigger a most relevant gene induction as proposed by Roberts is a true paradigm shift and also contradicting all previous teaching by the same author.

2. Stable biological endpoint

Currently, one key requirement for validation of *in vitro* methods is mechanistic relevance of reproducible *in vitro* endpoints. Two observations from the literature of the GARD™ developers should be discussed in this respect. The original assay was developed with the Affymetrix platform (Johansson et al., 2013), later the approach was tested on different platforms (Forreryd et al., 2014) and finally it was transferred to the nanoString® nCounter technology (Forreryd et al., 2016). It was observed that this change of technology platform led to only moderate to poor correlation of the original transcriptomic signature to the new platform (Forreryd et al., 2014), yet the new platform again gave a similarly high predictivity. Thus the measured signature seems not to be a fixed biological endpoint measurable with different technical means – but it is partly dependent on detection technology. This is compensated for, obviously, by partial redundancy in information content in the gene signature which keeps reported predictivity high, independent of the exact molecular endpoints measured (Forreryd et al., 2016). However, it indicates that we do not deal here with a completely defined molecular endpoint. Secondly, it is described that there are batch effects in the GARD™ assay, necessitating that each test series includes a benchmarking set of 10 chemicals for fine-tuning of the prediction signature to correct for batch effects (Forreryd et al., 2016). This is different from other *in vitro* assays, where the prediction model is fixed and positive controls are only run for assay performance check.

3. Predictivity

The paper by Roberts makes strong arguments based on the superior predictivity of the published dataset vs. the OECD validated assays and

DOI of original article: <https://doi.org/10.1016/j.yrtph.2019.01.006>

<https://doi.org/10.1016/j.yrtph.2019.01.005>

Received 29 October 2018; Accepted 2 January 2019

Available online 07 January 2019

0273-2300/ © 2019 Elsevier Inc. All rights reserved.

vs. the simplest WoE approach, namely the ‘2 out of 3’ approach which is listed as an OECD case study (Guidance Document on the, 2016). One may, from a regulatory toxicology viewpoint, indeed claim that ‘all is well, that predicts well’. However this will require more independent evaluation. First, it should be noted, that the analysis of Roberts is comparing *in vitro* data to the LLNA and not to human data (albeit these are available for skin sensitization).

Comparing GARD™ results to LLNA data, Roberts is overlooking uncertainties arising from the choice of the data base and the limited precision of the assay. In an independent Cosmetic Europe evaluation with blind coded chemicals (Johansson et al., 2017), the balanced accuracy was at 75%, while it was 88% in the in-house validation for the GARD assay. This lower balanced accuracy may be partly due to a bias of a low number of non-sensitizers tested in the Cosmetics Europe study. Nevertheless, a similar effect was noted by Cosmetic Europe when evaluating another assay based on multiple transcripts, where the balanced accuracy for the independent evaluation was 19% lower than the one reported from an in-house dataset, an effect which was not seen to that extend for the OECD validated assays. In addition to the chosen dataset, also limited precision of testing methods results in uncertainties of their predictive accuracy metrics. Hence the comparison of distinct numbers – Roberts compares 88% accuracy of the GARD™ assay to 83% of the “2 out of 3” approach – is neglecting this uncertainty. The 95% confidence interval of the accuracy of the “2 out of 3” approach is 74–97%; uncertainty information on the GARD™ assay is missing, but there is clearly an overlap rather than a significant difference in the accuracy measures.

4. Lack of transparency

It was often stated that the Tox21 approach will lead us into a mechanistic toxicology, which brings us away from the ‘black box’ animal test. The purely statistical approach of a transcriptomic analysis on > 200 transcripts (which are dependent on the detection methodology), selected from a pan-genomic analysis, however leads to a black-box outcome again, and there is no full transparency on what is determining the predictions being made, although some interrogation of relevant pathways was made.

5. Gard potency – predicting potency without dose-response information

The mechanistically most challenging proposal from the GARD platform is the recently published potency assay (Zeller et al., 2017), which claims a good discrimination of GHS sub-categories 1A and 1B by differential gene expression. The GARD assay does make a concentration range finder, but then test chemicals at only one defined concentration at 90% cellular viability. As described in the paper, the selected test concentration does not form part of the prediction model, which is entirely based on a gene signature and does not include dose-response information. It is of note, that these two GHS sub-categories (GHS sub-category 1A assigned for high incidence in humans or high potency as indicated by LLNA EC₃ < 2%) are based on animal test dose-response data and the threshold of EC₃ < 2% is purely a regulatory decision without additional biological meaning. It would be logical, that this arbitrary threshold in a continuous dose-dependent variable in an animal test would best be modeled based on dose-response

information from an *in vitro* test or a battery of *in vitro* tests. Although the authors have done a careful discrimination between training and test set, it is intriguing and biologically not intuitive that it is possible to classify chemicals into two regulatory defined potency sub-categories based on a gene signature not taking dose-response information into account. It will be very important to safeguard against overfit when making such proposals.

In conclusion, the strong proposals made in the paper by Roberts and by the GARD developers raise several questions which need careful discussion by the toxicological community before we can confirm the superior predictivity of the statistics-derived single cell-type method and move away from the AOP paradigm. Statistics-derived methods in toxicology may harness ‘big data’ and hence be information-rich but also lose on the transparency and biological as well as on the mechanistic side.

Transparency document

Transparency document related to this article can be found online at <https://doi.org/10.1016/j.yrtph.2019.01.005>.

References

- Forreryd, A., Johansson, H., Albrekt, A.S., Lindstedt, M., 2014. Evaluation of high throughput gene expression platforms using a genomic biomarker signature for prediction of skin sensitization. *BMC Genomics* 15, 379.
- Forreryd, A., Zeller, K.S., Lindberg, T., Johansson, H., Lindstedt, M., 2016. From genome-wide arrays to tailor-made biomarker readout - progress towards routine analysis of skin sensitizing chemicals with GARD. *Toxicol. Vitro* 37, 178–188.
- Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to Be Used within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation. ENV/JM/MONO, pp. 29. <https://doi.org/10.1787/9789264279285-en>.
- Johansson, H., Gradin, R., 2017. Skin sensitization: challenging the conventional thinking-A case against 2 out of 3 as integrated testing strategy. *Toxicol. Sci.* 159, 3–5.
- Johansson, H., Albrekt, A.S., Borrebaeck, C.A., Lindstedt, M., 2013 Apr. The GARD assay for assessment of chemical skin sensitizers. *Toxicol. Vitro* 27 (3). <https://doi.org/10.1016/j.tiv.2012.05.019>.
- Johansson, H., Gradin, R., Forreryd, A., Agemark, M., Zeller, K., Johansson, A., Larne, O., van Vliet, E., Borrebaeck, C., Lindstedt, M., 2017. Evaluation of the GARD assay in a blind Cosmetics Europe study. *ALTEX* 34, 515–523.
- Jowsey, I.R., Basketter, D.A., Westmoreland, C., Kimber, I., 2006. A future approach to measuring relative skin sensitising potency: a proposal. *J. Appl. Toxicol.* 26, 341–350.
- OECD, 2012. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins, Part 1: Scientific Evidence. OECD ENVIRONMENT, HEALTH AND SAFETY PUBLICATIONS SERIES ON TESTING AND ASSESSMENT NO. 168.
- Roberts, D.W., 2018. Is a combination of assays really needed for non-animal prediction of skin sensitization potential? Performance of the GARD (Genomic Allergen Rapid Detection) assay in comparison with OECD guideline assays alone and in combination. *Regul. Toxicol. Pharmacol.* 98, 155–160.
- Zeller, K.S., Forreryd, A., Lindberg, T., Gradin, R., Chawade, A., Lindstedt, M., 2017. The GARD platform for potency assessment of skin sensitizing chemicals. *ALTEX* 34, 539–559.

Andreas Natsch*

Givaudan Schweiz AG, Ueberlandstrasse 138, CH-8600, Duebendorf, Switzerland

E-mail address: andreas.natsch@givaudan.com.

Robert Landsiedel

BASF SE, RB/TB – Z470, D-67056, Ludwigshafen, Germany

* Corresponding author.