Research Article

Evaluation of the GARD Assay in a Blind Cosmetics Europe Study

Henrik Johansson [1], Robin Gradin [1], Andy Forreryd [2], Maria Agemark [1], Kathrin Zeller [2], Angelica Johansson [1], Olivia Larne [1], Erwin van Vliet [3], Carl Borrebaeck [2] and Malin Lindstedt [2]

[1]SenzaGen AB, Lund, Sweden; [2]Dept. of Immunotechnology, Lund University, Lund, Sweden; [3]Cosmetics Europe – The Personal Care Association, Brussels, Belgium

Summary

Chemical hypersensitivity is an immunological response towards foreign substances, commonly referred to as sensitizers, which gives rise primarily to the clinical symptoms known as allergic contact dermatitis. For the purpose of mitigating risks associated with consumer products, chemicals are screened for sensitizing effects. Historically, such predictive screenings have been performed using animal models. However, due to industrial and regulatory demand, animal models for the purpose of sensitization assessment are being replaced by non-animal testing methods, a global trend that is spreading across industries and market segments. To meet this demand, the Genomic Allergen Rapid Detection (GARD) assay was developed. GARD is a novel, cell-based assay that utilizes the innate recognition of xenobiotic substances by dendritic cells, as measured by a multivariate readout of genomic biomarkers. Following cellular stimulation, chemicals are classified as sensitizers or non-sensitizers based on induced transcriptional profiles. Recently, a number of non-animal methods were comparatively evaluated by Cosmetics Europe, using a coherent and blinded test panel of reference chemicals with human and local lymph node assay data, comprising a wide range of sensitizers and non-sensitizers. The outcome of the GARD assay is presented in this paper. It was demonstrated that GARD is a highly functional assay with a predictive performance of 83% in this Cosmetics Europe dataset. The average accumulated predictive accuracy of GARD across independent datasets was 86% for skin sensitization hazard.

Keywords: GARD, sensitization, *in vitro*, predictive accuracy, alternative methods

1 Introduction

Chemical hypersensitivity is a disease state induced by the human immune system in response to chemical sensitizers, which most frequently gives rise to the clinical symptoms of allergic contact dermatitis (ACD). The molecular and cellular mechanisms of sensitization have been reviewed extensively (Ainscough et al., 2013; Martin, 2015; Martin et al., 2011). Briefly, sensitization involves skin penetration of the sensitizing agent with a subsequent haptenization of endogenous proteins. Protein-hapten complexes are taken up by resident dendritic cells (DCs), which upon maturation migrate to local lymph nodes where antigen presentation to naïve T cells occurs. This results in the induction of an immunologic memory towards the specific sensitizer. Upon repeated exposure, a sensitized individual will suffer from ACD-associated symptoms following the elicitation of specific Th1 and cytotoxic $CD8^+$ T-cells.

A link has been made between the prevalence of ACD and the increased exposure of the population to the abundance of chemical sensitizers in consumer products (Lunder and Kansky, 2000; Nguyen et al., 2008). In order to limit hazardous effects of chemicals, risk assessments aim at safeguarding humans and the environment by eliminating and mitigating risks of exposure. The European REACH (EU, 2006) legislation requires all manufactured substances to undergo safety testing in order to identify, e.g., chemical sensitizers. Historically, such tests have been conducted in guinea pig (Magnusson and Kligman, 1969) and murine (Basketter et al., 2002) models. Mainly, the murine Local Lymph Node Assay (LLNA) continues to be used today. However, the use of animals for testing cosmetic ingredients has been banned in the EU since 2013 (EU, 2009), and the REACH legislation urges other industries to use animal testing only as a last resort when no relevant alternative testing methods exists, thereby clearly stating an intent to comply with the 3R principles (Russel and Burch, 1959).

Received January 12, 2017;
Accepted February 14, 2017;
Epub February 17, 2017;
doi:10.14573/altex.1701121

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.

As a consequence, the field of predictive toxicology has recently seen a surge in the development of novel non-animal assays for the assessment of chemical sensitization potential. The Direct Peptide Reactivity Assay (DPRA) (Gerberick et al., 2004), KeratinoSens™ (Natsch, 2010) and the human Cell Line Activation Test (h-CLAT) (Ashikaga et al., 2006) have been validated by the European Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM) and have recently been accepted by the OECD as test guidelines, which demonstrates that these tests are adequately reproducible and transferable (DPRA, OECD TG 442C; Keratinosens, OECD TG 442D; h-CLAT, OECD TG 442E).

However, none of the aforementioned assays are thought to fully cover the complexity of the skin sensitization process as stand-alone tests. Rather, it is widely proposed that assessment of hazard and/or risk should be carried out using integrated testing strategies (ITS), also referred to as integrated approaches to testing and assessment (IATA) (Jaworska and Hoffmann, 2010; Hartung et al., 2013; Rovida et al., 2015; Ezendam et al., 2016). However, the overall predictive performance of an ITS will invariably depend on the predictivity of its assay constituents. In addition, being based on a single or a few biomarkers, current methods provide only limited predictive information, as well as sometimes overlapping mechanistic information. Thus, when designing an ITS, tests with high predictive performance and information content, covering one or more of the key events of the adverse outcome pathway (AOP) (OECD, 2012), would clearly be an advantageous option (Lindstedt and Borrebaeck, 2011).

The Genomic Allergen Rapid Detection (GARD) assay is a cell-based *in vitro* assay for assessment of chemical sensitizers (Johansson et al., 2011). The readout of the assay is based on differentially regulated transcriptional changes of selected genomic biomarkers, referred to as the GARD prediction signature (GPS), induced in a myeloid dendritic cell-like cell line in response to chemical stimulation. GARD has been shown to be functional and able to accurately predict sensitizing chemicals in blind evaluations (Johansson et al., 2014) and exhibits high predictive performance in comparison with *in vitro* counterparts (Johansson and Lindstedt, 2014). Following a thorough evaluation of technological platforms (Forreryd et al., 2014), the assay was recently adapted to a medium-to-high throughput format in order to meet industrial and regulatory demands of reliability, resource effectiveness and sample capacity (Forreryd et al., 2016). Furthermore, an adaptation of GARD using identical cellular protocols but a different biomarker signature to differentially classify respiratory sensitizers from a set of skin sensitizers and non-sensitizers has been demonstrated (Forreryd et al., 2015). This illustrates the unparalleled flexibility of applications of genomics-based platforms, which is due to the massive amount of information that multivariate readouts deliver.

In an attempt to evaluate the performance of currently validated assays, as well as selected assays that are currently in the validation process or are being considered for validation, the Cosmetics Europe Skin Tolerance Task Force (CE STTF) recently published a comparative study in which a limited set of chemicals were classified as sensitizers or non-sensitizers (Reisinger et al., 2015). Based on this study, the best-performing assays, among them GARD, were selected for a second evaluation phase comprising a larger number of blinded chemicals with human and LLNA data. Here, we report the predictive performance of GARD on this Cosmetics Europe dataset as well as an updated overall predictive accuracy of the assay, calculated using strictly independent sets of test chemicals.

2 Materials and methods

Chemicals and datasets

A dataset for model training, consisting of 40 different cell stimulations in biological triplicates, was defined previously and the respective dataset details are described elsewhere (Johansson et al., 2011; Forreryd et al., 2016). In this study, a total of 73 chemicals (see Tab. 1 for details) were assayed blindly using the above-mentioned training data set. All chemicals were provided by the CE STTF, which also kept the code for the blinded chemicals. All chemicals were stored according to the suppliers' recommendations. In addition to the blinded chemicals of the test set, a set of non-blind benchmark controls (see Tab. S1 at doi:10.14573/altex.1701121s for details) were included. The purpose of the benchmark controls was to calibrate the prediction model to the current batch of cells, as described (Forreryd et al., 2016). All chemicals used as benchmark controls were purchased from Sigma Aldrich (St. Louis, MO, USA) and were stored according to the manufacturer's instructions.

Tab. 1: Dataset details and test results

Chemical identifiers		References			Assay parameters				GARD output		
Substance ID	CAS	LLNA	HP	GHS/CLP	vehicle	c.max	c.rv90	c.input	cDV (± SD)		Prediction
Sensitizers											
1,4-phenylenediamine	106-50-3	strong	1	1A	DMSO	500	70	70	5.8	± 0.6	S
Tetrachlorosalicylanilide	1154-59-2	extreme	1	–	DMSO	500	20	20	3.3	± 0.2	S
Dimethyl fumarate	624-49-7	strong	1	–	DMSO	500	–	90	5.9	± 0.4	S

Chemical identifiers		References			Assay parameters				GARD output		
Substance ID	CAS	LLNA	HP	GHS/ CLP	vehicle	c.max	c.rv90	c.input	cDV (± SD)		Prediction
2-aminophenol	95-55-6	strong	2	1A	DMSO	500	80	80	6.1	± 1.6	S
2-Nitro-1,4-phenylenediamine	5307-14-2	moderate	2	1A	DMSO	500	200	200	3.6	± 0.5	S
Formaldehyde (act.. 37%)	50-00-0	strong	2	1A	DMSO	500	260	260	1.2	± 0.5	S
Glutaraldehyde (act. 50%)	111-30-8	extreme	2	1A	DMSO	500	100	100	2.8	± 2.2	S
Methyl heptine carbonate	111-12-6	strong	2	1A	DMSO	500	50	50	0.2	± 0.8	S
Propyl gallate	121-79-9	strong	2	1A	DMSO	500	100	100	6.7	± 1.9	S
Toluene diamine sulphate	615-50-9	strong	2	–	DMSO	500	100	100	1.9	± 1	S
Glyoxal (act. 40%)	107-22-2	strong	2	1A	DMSO	500	–	500	0.8	± 1.1	S
Isoeugenol	97-54-1	moderate	2	1A	DMSO	500	500	500	6	± 0.4	S
1,2-Benzisothiazolin-3-one	2634-33-5	moderate	2	–	DMSO	500	12.5	12.5	1.6	± 0.7	S
3-dimethylaminopropylamine	109-55-7	moderate	2	–	DMSO	500	–	500	0.3	± 0.2	S
Thioglycerol	96-27-5	moderate	2	–	DMSO	500	–	500	-0.8	± 0.5	NS
Lyral	31906-04-4	weak	2	1B	DMSO	400	200	200	2.9	± 0.6	S
Chlorpromazine	50-53-3	moderate	3	1A	DMSO	100	10	10	1.9	± 0.9	S
Benzoyl peroxide	94-36-0	extreme	3	–	DMSO	500	85	85	-1.1	± 0.4	NS
Bisphenol A-diglycidyl ether	1675-54-3	moderate	3	1A	DMSO	200	50	50	3.4	± 1.9	S
Ethylene diamine	107-15-3	moderate	3	1B	DMSO	500	–	500	1.4	± 2.5	S
Glyceryl monothioglycolate	30618-84-9	moderate	3	1B	DMSO	500	200	200	1.2	± 1.5	S
Farnesol	4602-84-0	moderate	3	–	DMSO	500	–	500	2.1	± 0.9	S
Abietic acid	514-10-3	weak	3	1B	DMSO	200	–	200	1.3	± 1.3	S
Butyl glycidyl ether	2426-08-6	weak	3	1B	DMSO	500	480	480	3.6	± 2.1	S
Cinnamic alcohol	104-54-1	weak	3	1B	DMSO	500	–	500	8.8	± 1	S
Citral	5392-40-5	moderate	3	1B	DMSO	500	80	80	5.8	± 0.7	S
Eugenol	97-53-0	weak	3	1B	DMSO	500	400	400	2.6	± 0.4	S
Imidazolidinyl urea	39236-46-9	weak	3	1B	dH2O	500	50	50	1.5	± 2.4	S
Penicillin G	61-33-6	weak	3	–	DMSO	500	–	500	-0.8	± 0.9	NS
5-methyl-2,3-hexanedione	13706-86-0	weak	3	–	DMSO	500	–	500	3.1	± 0.9	S
Coumarin	91-64-5	NS	3	–	DMSO	500	–	500	0.3	± 0.8	S
Hexyl salicylate	6259-76-3	strong	4	1A	DMSO	500	120	120	-0.9	± 0.1	NS
Iodopropynyl butylcarbamate	55406-53-6	strong	4	1A	DMSO	500	10	10	0.7	± 1.5	S
Neomycin sulphate	1405-10-3	NS	4	–	dH2O	500	–	500	0.7	± 2	S
Resorcinol	108-46-3	moderate	4	1B	dH2O	500	–	500	2	± 1.1	S
Amylcinnamyl alcohol	101-85-9	NS	4	1B	DMSO	500	260	260	2.1	± 1.3	S
Aniline	62-53-3	weak	4	1B	DMSO	500	–	500	0.4	± 2.2	S
Benzocaine	94-09-7	NS	4	1B	DMSO	500	–	500	0.8	± 1.7	S
Geraniol	106-24-1	weak	4	1B	DMSO	500	–	500	2.4	± 1.7	S
Lillial	80-54-6	weak	4	1B	DMSO	500	160	160	1.7	± 0.5	S
Linalool	78-70-6	weak	4	1B	DMSO	500	–	500	0.6	± 0.8	S

Chemical identifiers		References			Assay parameters				GARD output		
Substance ID	**CAS**	**LLNA**	**HP**	**GHS/ CLP**	**vehicle**	**c.max**	**c.rv90**	**c.input**	**cDV (± SD)**		**Prediction**
Amyl cinnamic aldehyde	122-40-7	weak	4	–	DMSO	500	110	110	5.3	± 1.2	S
Carvone	6485-40-1	weak	4	–	DMSO	500	–	500	2.3	± 0.7	S
Kanamycin	70560-51-9	NS	4	–	dH2O	200	–	200	0.2	± 1.2	S
Anethole	104-46-1	moderate	5	1B	DMSO	500	–	500	2.3	± 1.5	S
Anisyl alcohol	105-13-5	moderate	5	1B	DMSO	500	–	500	0.1	± 1.5	S
Benzyl salicylate	118-58-1	moderate	5	–	DMSO	500	200	200	0.6	± 1.4	S
Limonene	5989-27-5	weak	5	1B	DMSO	500	–	500	0	± 0.4	S
Hexyl cinnamic aldehyde	101-86-0	weak	5	1B	DMSO	500	100	100	1.1	± 0.8	S
Benzyl benzoate	120-51-4	weak	5	1B	DMSO	500	500	500	2.3	± 1.8	S
Citronellol	106-22-9	weak	5	1B	DMSO	500	–	500	1.8	± 0.7	S
Diethanolamine	111-42-2	weak	5	1B	DMSO	500	–	500	0.5	± 0	S
Pentachlorophenol	87-86-5	weak	5	1B	DMSO	200	150	150	3.1	± 0.8	S
Pyridine	110-86-1	weak	5	1B	DMSO	500	–	500	0.4	± 0.2	S
Non-sensitizers											
Hydrocortisone	50-23-7	NS	5	no cat.	DMSO	500	–	500	5.9	± 0.1	S
Isopropanol	67-63-0	NS	5	no cat.	DMSO	500	–	500	-0.9	± 0.8	NS
Methyl salicylate	119-36-8	NS	5	no cat.	DMSO	500	–	500	0.2	± 2.4	S
Phenoxyethanol	122-99-6	NS	5	no cat.	DMSO	500	–	500	-0.3	± 1.3	NS
Propylene glycol	57-55-6	NS	5	no cat.	DMSO	500	–	500	-1.3	± 0.7	NS
Triethanolamine	102-71-6	NS	5	–	DMSO	500	–	500	0.2	± 2.4	S
4-aminobenzoic acid	150-13-0	NS	5	no cat.	DMSO	500	–	500	-1.4	± 0.8	NS
Benzaldehyde	100-52-7	NS	5	no cat.	DMSO	500	–	500	0	± 1.2	NS
Propyl paraben	94-13-3	NS	5	–	DMSO	500	–	500	5.3	± 0.1	S
Vanillin	121-33-5	NS	5	no cat.	DMSO	500	–	500	-2.4	± 0.9	NS
Dextran	9004-54-0	NS	6	no cat.	DMSO	40	–	40	-1	± 0.5	NS
Glycerol/Glycerin	56-81-5	NS	6	no cat.	DMSO	500	–	500	-0.5	± 0.8	NS
Octanoic acid	124-07-2	NS	6	no cat.	DMSO	500	–	500	-0.2	± 1.1	NS
Phenol	108-95-2	NS	6	no cat.	DMSO	500	–	500	-0.3	± 1.9	NS
Tocopherol	59-02-9	moderate	6	–	DMSO	100	–	100	0.7	± 1.7	S
Diethyl phthalate	84-66-2	NS	6	no cat.	DMSO	500	–	500	1.9	± 1	S
Diethyl toluamide	134-62-3	NS	6	–	DMSO	500	–	500	1.3	± 0.4	S
Tween 80	9005-65-6	NS	6	no cat.	DMSO	500	13	13	1.9	± 1.4	S

LLNA, Local Lymph Node Assay (as listed in the CE STTF database); HP, human potency (as listed in Basketter et al., 2014);
GHS/CLP, Global Harmonization System / Classification for Labelling and Packaging (as listed in Piroird et al., 2015);
c.max, maximum concentration of titration range (µM); rv90, in-well concentration yielding 90% relative viability (µM); c.input, concentration used for cell stimulation, derived from c.max and c.rv90 (µM); c.DV, calibrated decision value; NS, non-sensitizer; S, sensitizer. For details, see Johansson et al., 2013.

Cell maintenance, chemical stimulations, phenotypic analysis and total RNA isolation

All GARD protocols for cell maintenance, cellular stimulation with chemicals, required phenotypical quality control of cells prior to chemical stimulation, and isolation of total RNA have been described previously (Johansson et al., 2013, 2011; Forreryd et al., 2016) and were followed without deviation in this study. The myeloid cell line used in this study was derived from MUTZ-3 (DSMZ, Braunschweig, Germany) and is available via SenzaGen AB (SenzaGen AB, Lund, Sweden). All cellular stimulations were performed in biological triplicates, using separate cell batches for each replicate. Following chemical stimulation, cells were harvested and lysed with TRizol reagent (Thermo Scientific, Waltham, MA), and stored at -20°C until RNA extraction. Total RNA was isolated from lysed samples using Direct-zol™ RNA MiniPrep column purification kit (Zymo Research, Irvine, CA, USA) according to protocols provided by the manufacturer. Total RNA concentrations and RNA integrity were assessed using the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Total RNA was stored at -80°C until NanoString nCounter analysis.

Gene expression analysis using NanoString technology

The design of a custom NanoString CodeSet, corresponding to the GARD prediction signature (GPS), was described recently (Forreryd et al., 2016). All NanoString-associated protocols for gene expression analysis were performed according to instructions by the manufacturer. In short, the custom CodeSet was hybridized with 100 ng total RNA (5 µl at 20 ng/µl) and incubated at 65°C for 24 h. Hybridized samples were processed in the NanoString GEN2 nCounter Prep Station 5s, using the High Sensitivity protocol, and analyzed in the NanoString Digital Analyzer 5s for digital quantification of each transcript of the GPS, using maximal resolution (555 fields of view). All required equipment, CodeSet and master kit reagents were obtained from NanoString Technologies (NanoString Technologies, Seattle, WA, USA).

Data pre-processing, normalization and analysis

Raw nCounter gene expression data was imported into the R statistical environment (R Development Core Team, 2014), in which all downstream analysis was performed. Data was normalized using a counts per total counts (CPTC) algorithm, which reports normalized values for any given gene of the GPS as the ratio of digital counts for the specific gene and the total counts of all measured genes within that sample. Generation of prediction calls for each sample (sensitizer/non-sensitizer) was performed as described previously. Briefly, a support vector machine (SVM) (Cortes and Vapnik, 1995) was trained on the training dataset and used to generate decision values (DVs) for each sample of the benchmark control dataset and test dataset, respectively. The predictive performance of the model was evaluated on the benchmark control dataset using the additional R package ROCR (Sing et al., 2005). Observations of the receiver operating characteristic (ROC) (Lasko et al., 2005) allowed the identification of the prediction model cutoff that achieves the highest accuracy of predictions of the benchmark control dataset, which was subsequently subtracted from all DVs generated

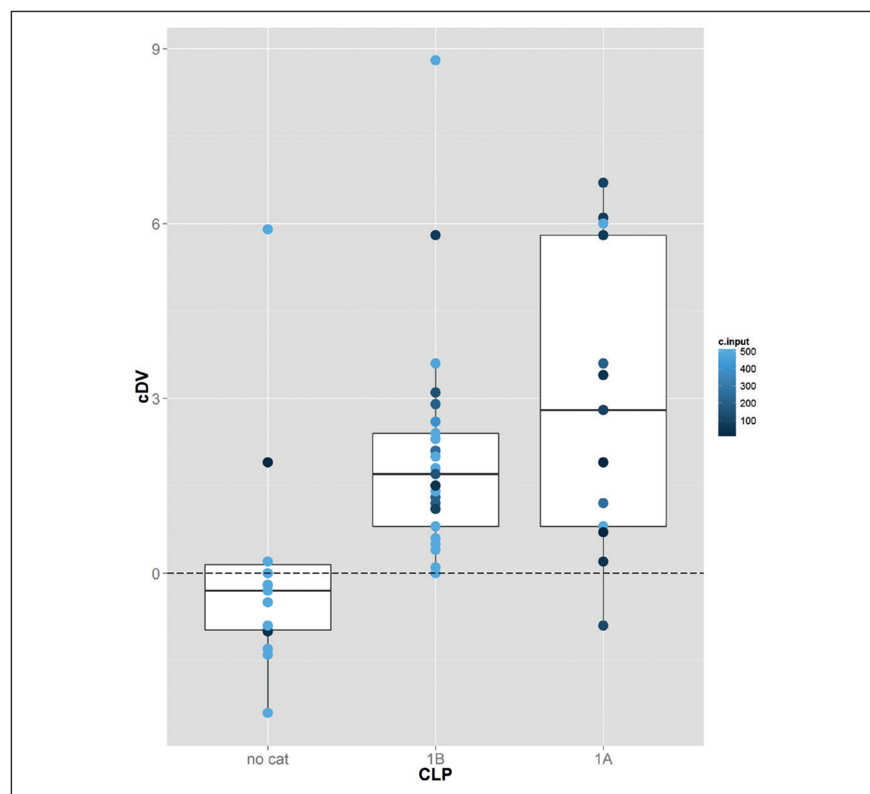

Fig. 1: GARD predictions correlate with potency classifications

Box-and-whisker plots of mean GARD cDVs, grouped by sensitizing potency as defined by the GHS/CLP classification system. Only test substances for which such classifications are available are included, see Table 1 (n chemicals = 52). The color of each data point is mapped to the GARD input concentration (µM) used for that test substance.

from samples of the test dataset. Thus, final predictions were performed on calibrated DVs (cDVs). A specific chemical used for stimulation was classified as a sensitizer if the mean cDV from biological triplicates was greater than zero. The predictive performance of the model's classifications of the test dataset was assessed using Cooper statistics (Cooper et al., 1979).

3 Results

3.1 GARD classifications of the blinded CE-reference panel of chemicals

A set of blinded chemicals was classified as sensitizers or non-sensitizers by the GARD assay using established protocols. GARD predictions of the chemicals used in this study are presented in Table 1. Calculations of various predictive performance parameters based on Cooper statistics are presented in Table 2. For the purpose of binary predictions, a composite reference was defined to classify a sensitizer as a compound that is categorized as having a human potency (HP) (Basketter et al., 2014) of 1-4, or being categorized as HP 5, if it is also predicted as a sensitizer by the LLNA. Consequently, compounds categorized as HP 5, predicted as non-sensitizers by the LLNA, are here defined as non-sensitizers, together with all compounds of HP 6. This binary classification system perfectly correlates with the Global Harmonization System (GHS) / Classification for Labelling and Packaging (CLP) classifications. By this definition, based on the current data, the accuracy, specificity and sensitivity of GARD, is 83%, 56% and 93%, respectively. Comparing GARD predictions strictly with either HP or LLNA, the concordance was estimated to be 81% and 76%, respectively. The mean magnitudes of the cDVs are visualized in box-and-whisker plots in Figure 1, grouped according to their sensitizing potency as defined by the GHS/CLP system. The observed differences in mean cDVs indicate that the GARD predictions correlate with potency classifications.

Tab. 2: Cooper statistics of current data

Characteristic	LLNA	HP	Composite
Accuracy (%)	76	81	83
Sensitivity (%)	90	84	93
Specificity (%)	45	50	56

LLNA, Local Lymph Node Assay; HP, human potency

3.2 Accumulated GARD performance parameters across historical datasets

In order to relate the current results to previously published figures of predictive performance, an update of accumulated Cooper statistics for independent GARD assessments across various datasets is presented in Table 3. Combining datasets from a total of 127 chemicals, the accuracy of GARD was calculated to be 86%.

4 Discussion

In the last decade, substantial efforts have been made to develop and validate alternative non-animal assays for the assessment of chemical sensitizers in order to meet changing regulatory and industrial demands. The current leading opinion is that no single assay is likely to provide sufficient information for accurate safety assessment of chemicals as a stand-alone test. This notion is supported by the data generated by currently validated tests and the subsequent recommendations given by EURL ECVAM (EC, 2013, 2014, 2015). For this reason, it is of great importance to continuously compare and evaluate novel and already established test methods using coherent reference chemical panels in order to prioritize assays that display superior functionality and predictivity when designing IATAs, or in the quest for stand-alone tests.

In this report, we present novel data regarding the functionality and predictive performance of GARD, generated in a blind study performed in association with the CE STTF. In this independent dataset, GARD accurately classified 83% out of a total of 72 chemicals for skin sensitization hazard. Adding this figure to previously published data from independent evaluation studies, GARD displays an accumulated accuracy of 86%, based on the classification of a total of 127 chemicals.

It is appropriate at this point to consider the gold standard of sensitization assessment, i.e., the reference against which such performance estimations are calculated. In this report, comparisons have been made with both LLNA classifications and human potency (HP), as defined by Basketter et al. (2014). The concordance of GARD with these metrics was 76% and 81%, respectively. Of note, the concordance between LLNA and

Tab. 3: Accumulated predictive performance

Dataset	Sensitivity		Specificity		Accuracy		Source
GARD in-house validation	89%	(17/19)	86%	(6/7)	88%	(23/26)	Johansson, 2014
Technology transfer and method optimization	94%	(16/17)	83%	(10/12)	90%	(26/29)	Forreryd, 2016
Current study	93%	(50/54)	56%	(10/18)	83%	(60/72)	–
Accumulated predictive performance	92%	(83/90)	70%	(26/37)	86%	(109/127)	–

HP within the same data is 78%, clearly demonstrating that perfect correlation with either metric is mutually exclusive. In particular, HP category 5 includes numerous compounds that have historically been classified as both sensitizers (e.g., hexyl cinnamic aldehyde) and non-sensitizers (e.g., isopropanol). For this reason, a composite reference was proposed for binary classifications, in which a sensitizer was defined to include HP categories 1-4, together with chemicals of the HP category 5, for which the LLNA classification was positive. Still, looking at the present data, we find that GARD misclassifications are overrepresented in HP category 5. Considering only chemicals assigned within the HP categories 1-4, GARD accurately predicts 91% as sensitizers, while the corresponding accuracy within category 5 is 80%. Based on the reasoning above, it is logical to assume that the annotations provided as a reference may include errors based on flawed conclusions, as discussed (Basketter et al., 2014).

On a chemical by chemical basis, false GARD classifications were obtained for thioglycerol, benzoyl peroxide, penicillin G, hexyl salicylate (false negatives, HP category 1-4), hydrocortisone, methyl salicylate, triethanolamine, propyl paraben (false positives, HP category 5) and tocopherol, diethyl phthalate, diethyl toluamide and Tween 80 (false positives, HP category 6).

For the false negatives, the obvious common denominator is that a majority fails to induce any cytotoxic effect in the present cellular system. It should be noted, however, that inducing cytotoxic effects is not a requirement for the successful assessment of a sensitizer. Indeed, numerous examples of correctly classified sensitizers that do not induce cytotoxicity are available within this dataset. Correspondingly, toxic effects are not exclusively induced by sensitizers. It has previously been observed that non-toxic compounds are overrepresented among false negatives (Johansson et al., 2014). Furthermore, the connection between toxic or irritating effects and induction of sensitization has previously been discussed by other authors (Nukada et al., 2011). Thus, the overrepresentation of misclassifications among non-toxic sensitizers is a problem shared with many cell-based assays. For false negatives that do induce cytotoxicity, no apparent explanation is available at this point.

The false positives among HP category 5 are, as discussed above, likely related to the ambiguous annotations provided by current gold standards. Indeed, the fact that they are listed within HP category 5 separates them from true non-sensitizers, at least by one metric, suggesting that observed LLNA classifications are non-concordant with the effects observed in the clinic. As an example, clinical cases of sensitization towards hydrocortisone are indeed not infrequent (Burden and Beck, 1992). Thus, the correctness of calling such substances non-sensitizers, and thereby concluding that GARD produces misclassifications, is certainly controversial.

Finally, false positives within HP category 6 include tocopherol, which is classified as a moderate sensitizer by the LLNA. Furthermore, diethyl phthalate and diethyl toluamide are both frequently classified as positives in cell-based assays (Ashikaga et al., 2010; Piroird et al., 2015), while Tween 80 is consistently classified as a sensitizer in numerous assays (Emter et al., 2013; Piroird et al., 2015; Ramirez et al., 2014). Indeed, the sensitizing capacity of Tween 80 has been closely examined and confirmed to be evident both before and after oxidation (Bergh et al., 1997). Consequently, the inherent difficulty of accurately assessing these compounds should rather be regarded as general. Naturally, these aspects were a contributing factor to including such compounds in the blinded dataset used in this study, likely skewing the estimated specificity within the dataset towards lower figures compared to what would be expected in broader chemical domains.

During GARD development, it was observed that the relative magnitude of the GARD decision values correlates with sensitizing potency (Johansson et al., 2011), a hypothesis that has been maintained since. In light of the above discussed ambiguities regarding sensitizing potency, as estimated by current gold standards, GARD development towards potency assessment focuses on the distinction between strong and weak sensitizers in accordance with the GHS/CLP classification system. In Figure 1, the cDVs of the test substances are grouped according to this system. From the current data it is clear that the hypothesis based on earlier observations prevails, since strong sensitizers (1A) on average generate higher DVs compared to weak sensitizers (1B). Furthermore, it is evident that the cytotoxicity of a chemical is also related to its sensitizing potency. In current GARD protocols, cytotoxic compounds are used at concentrations that maintain 90% relative cell viability. From Figure 1, it is evident that strong sensitizers (1A) are on average assayed at lower concentrations compared to weak sensitizers (1B), due to their higher levels of cytotoxic effects. While the GARD platform indeed holds information regarding sensitizing potency, there is an overlap between the different categories, which presently hampers its utilization for accurate potency assessment. However, the harnessing of accurate potency information is currently being refined for accurate sub-categorization (manuscript in preparation).

In conclusion, we here report data of GARD performance on an extended, blinded set of chemicals. Taken together, GARD is consistently functional across datasets, with a predictive accuracy of 83% in this Cosmetics Europe dataset and average predictive accuracy of 86% in a combined dataset of 127 chemicals for skin sensitization hazard.

References

Ainscough, J. S., Gerberick, G. F., Dearman, R. J. and Kimber, I. (2013). Danger, intracellular signaling, and the orchestration of dendritic cell function in skin sensitization. *J Immunotoxicol 10*, 223-234. doi:10.3109/1547691X.2012.711782

Ashikaga, T., Yoshida, Y., Hirota, M. et al. (2006). Development of an in vitro skin sensitization test using human cell lines: The human cell line activation test (h-CLAT). I. Optimization of the h-CLAT protocol. *Toxicol In Vitro 20*, 767-773. doi:10.1016/j.tiv.2005.10.012

Ashikaga, T., Sakaguchi, H., Sono, S. et al. (2010). A comparative evaluation of in vitro skin sensitisation tests: The human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). *Altern Lab Anim 38*, 275-284.

Basketter, D. A., Evans, P., Fielder, R. J. et al. (2002). Local lymph node assay – Validation, conduct and use in practice. *Food Chem Toxicol 40*, 593-598. doi:10.1016/S0278-6915(01)00130-2

Basketter, D. A., Alepee, N., Ashikaga, T. et al. (2014). Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis 25*, 11-21. doi:10.1097/DER.0000000000000003

Bergh, M., Magnusson, K., Nilsson, J. L. and Karlberg, A. T. (1997). Contact allergenic activity of Tween 80 before and after air exposure. *Contact Dermatitis 37*, 9-18. doi:10.1111/j.1600-0536.1997.tb00368.x

Burden, A. D. and Beck, M. H. (1992). Contact hypersensitivity to topical corticosteroids. *Br J Dermatol 127*, 497-500. doi:10.1111/j.1365-2133.1992.tb14847.x

Cooper, J. A., 2nd, Saracci, R. and Cole, P. (1979). Describing the validity of carcinogen screening tests. *Br J Cancer 39*, 87-89. doi:10.1038/bjc.1979.10

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning 20*, 273-297. doi:10.1007/BF00994018

EC (2013). EURL ECVAM Recommendation on the Direct Peptide Reactivity Assay (DPRA) for Skin Sensitisation Testing. https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations/files-dpra/EURL_ECVAM_Recommendation_DPRA_2013.pdf

EC (2014). EURL ECVAM Recommendation on the KeratinoSensTM assay for skin sensitization testing. http://publications.jrc.ec.europa.eu/repository/bitstream/JRC87551/lbna26427enn.pdf

EC (2015). EURL ECVAM Recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitisation testing. https://eurl-ecvam.jrc.ec.europa.eu/news/news_docs/eurl-ecvam-recommendation-on-the-human-cell-line-activation-test-h-clat-for-skin-sensitisation-testing

Emter, R., van der Veen, J. W., Adamson, G. et al. (2013). Gene expression changes induced by skin sensitizers in the KeratinoSens cell line: Discriminating Nrf2-dependent and Nrf2-independent events. *Toxicol In Vitro 27*, 2225-2232. doi:10.1016/j.tiv.2013.09.009

EU (2006). Regulation (EC) No 1907/2006 of the European Parliament and of theCouncil of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Off J Eur Union L 396*, 1-1355. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02006R1907-20140410&from=EN

EU (2009). Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. *Off J Eur Union L342*, 1-59. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009R1223&from=EN

Ezendam, J., Braakhuis, H. M. and Vandebriel, R. J. (2016). State of the art in non-animal approaches for skin sensitization testing: From individual test methods towards testing strategies. *Arch Toxicol 90*, 2861-2883. doi:10.1007/s00204-016-1842-4

Forreryd, A., Johansson, H., Albrekt, A. S. and Lindstedt, M. (2014). Evaluation of high throughput gene expression platforms using a genomic biomarker signature for prediction of skin sensitization. *BMC Genomics 15*, 379. doi:10.1186/1471-2164-15-379

Forreryd, A., Johansson, H., Albrekt, A. S. et al. (2015). Prediction of chemical respiratory sensitizers using GARD, a novel in vitro assay based on a genomic biomarker signature. *PLoS One 10*, e0118808. doi:10.1371/journal.pone.0118808

Forreryd, A., Zeller, K. S., Lindberg, T. et al. (2016). From genome-wide arrays to tailor-made biomarker readout – Progress towards routine analysis of skin sensitizing chemicals with GARD. *Toxicol In Vitro 37,* 178-188. doi:10.1016/j.tiv.2016.09.013

Gerberick, G. F., Vassallo, J. D., Bailey, R. E. et al. (2004). Development of a peptide reactivity assay for screening contact allergens. *Toxicol Sci 81*, 332-343. doi:10.1093/toxsci/kfh213

Hartung, T., Luechtefeld, T., Maertens, A. and Kleensang, A. (2013). Integrated testing strategies for safety assessments. *ALTEX 30*, 3-18. doi:10.14573/altex.2013.1.003

Jaworska, J. and Hoffmann, S. (2010). Integrated Testing Strategy (ITS) – Opportunities to better use existing data and guide future testing in toxicology. *ALTEX 27*, 231-242. doi:10.14573/altex.2010.4.231

Johansson, H., Lindstedt, M., Albrekt, A. S. and Borrebaeck, C. A. (2011). A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. *BMC Genomics 12*, 399. doi:10.1186/1471-2164-12-399

Johansson, H., Albrekt, A. S., Borrebaeck, C. A. and Lindstedt, M. (2013). The GARD assay for assessment of chemical skin sensitizers. *Toxicol In Vitro 27*, 1163-1169. doi:10.1016/j.tiv.2012.05.019

Johansson, H. and Lindstedt, M. (2014). Prediction of skin sensitizers using alternative methods to animal experimentation. *Basic Clin Pharmacol Toxicol 115*, 110-117. doi:10.1111/bcpt.12199

Johansson, H., Rydnert, F., Kuhnl, J. et al. (2014). Genomic allergen rapid detection in-house validation – A proof of concept. *Toxicol Sci 139*, 362-370. doi:10.1093/toxsci/kfu046

Lasko, T. A., Bhagwat, J. G., Zou, K. H. and Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform 38*, 404-415. doi:10.1016/j.jbi.2005.02.008

Lindstedt, M. and Borrebaeck, C. (2011). Pattern rules: Biomarker signatures for sensitization as an alternative to animal testing. *Biomark Med 5*, 809-811. doi:10.2217/bmm.11.82

Lunder, T. and Kansky, A. (2000). Increase in contact allergy to fragrances: Patch-test results 1989-1998. *Contact Dermatitis 43*, 107-109. doi:10.1034/j.1600-0536.2000.043002107.x

Magnusson, B. and Kligman, A. M. (1969). The identification

of contact allergens by animal assay. The guinea pig maximization test. *J Invest Dermatol 52*, 268-276. doi:10.1038/jid.1969.42

Martin, S. F., Esser, P. R., Weber, F. C. et al. (2011). Mechanisms of chemical-induced innate immunity in allergic contact dermatitis. *Allergy 66*, 1152-1163. doi:10.1111/j.1398-9995.2011.02652.x

Martin, S. F. (2015). New concepts in cutaneous allergy. *Contact Dermatitis 72*, 2-10. doi:10.1111/cod.12311

Natsch, A. (2010). The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers – Functional relevance and a hypothesis on innate reactions to skin sensitizers. *Toxicol Sci 113*, 284-292. doi:10.1093/toxsci/kfp228

Nguyen, S. H., Dang, T. P., MacPherson, C. et al. (2008). Prevalence of patch test results from 1970 to 2002 in a multi-centre population in North America (NACDG). *Contact Dermatitis 58*, 101-106. doi:10.1111/j.1600-0536.2007.01281.x

Nukada, Y., Ito, Y., Miyazawa, M. et al. (2011). The relationship between CD86 and CD54 protein expression and cytotoxicity following stimulation with contact allergen in THP-1 cells. *J Toxicol Sci 36*, 313-324. doi:10.2131/jts.36.313

OECD (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins, Part 1: Scientific Evidence. *Environment, Health and Safety Publications, Series on Testing and Assessment No. 168*. http://www.oecd.org/env/the-adverse-outcome-pathway-for-skin-sensitisation-initiated-by-covalent-binding-to-proteins-9789264221444-en.htm

Piroird, C., Ovigne, J. M., Rousset, F. et al. (2015). The myeloid U937 skin sensitization test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. *Toxicol In Vitro 29*, 901-916. doi:10.1016/j.tiv.2015.03.009

R Development Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. http://www.R-project.org

Ramirez, T., Mehling, A., Kolle, S. N. et al. (2014). LuSens: A keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification. *Toxicol In Vitro 28*, 1482-1497. doi:10.1016/j.tiv.2014.08.002

Reisinger, K., Hoffmann, S., Alepee, N. et al. (2015). Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. *Toxicol In Vitro 29*, 259-270. doi:10.1016/j.tiv.2014.10.018

Rovida, C., Alepee, N., Api, A. M. et al. (2015). Integrated Testing Strategies (ITS) for safety assessment. *ALTEX 32*, 25-40. doi:10.14573/altex.1411011

Russel, W. and Burch, R. (1959). *The Principles of Humane Experimental Technique*. http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc

Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics 21*, 3940-3941. doi:10.1093/bioinformatics/bti623

Conflict of interest
The authors are employed or collaborate with SenzaGen, a company which commercializes the GARD test.

Acknowledgements
The authors would like to thank the Cosmetics Europe Skin Tolerance Task Force for providing the chemicals of the test data set and ensuring the blind integrity of the study.

Correspondence to
Henrik Johansson, PhD
SenzaGen AB
Medicon Village
Scheelevägen 2
22381 Lund, Sweden
Phone: +46 704 492724
e-mail: henrik.johansson@senzagen.com