

Genomic Allergen Rapid Detection In-House Validation—A Proof of Concept

Henrik Johansson,* Frida Rydnert,* Jochen Kühnl,† Andreas Schepky,† Carl Borrebaeck,* and Malin Lindstedt*,¹

*Department of Immunotechnology, Lund University, 22381 Lund, Sweden; and †Department of Toxicology, Beiersdorf AG, 20245 Hamburg, Germany

¹To whom correspondence should be addressed at Lund University, Medicon Village 406, 22381 Lund, Sweden. Fax: +46 46 2224200; E-mail: malin.lindstedt@immun.lth.se.

Received December 9, 2013; accepted March 4, 2014

Chemical sensitization is an adverse immunologic response to chemical substances, inducing hypersensitivity in exposed individuals. Identifying chemical sensitizers is of great importance for chemical, pharmaceutical, and cosmetic industries, in order to prevent the use of sensitizers in consumer products. Historically, chemical sensitizers have been assessed mainly by *in vivo* methods, however, recently enforced European legislations urge and promote the development of animal-free test methods able to predict chemical sensitizers. Recently, we presented a predictive biomarker signature in the myeloid cell line MUTZ-3, for assessment of skin sensitizers. The identified genomic biomarkers were found to be involved in immunologically relevant pathways, induced by recognition of foreign substances and regulating dendritic cell maturation and cytoprotective mechanisms. We have developed the usage of this biomarker signature into a novel *in vitro* assay for assessment of chemical sensitizers, called Genomic Allergen Rapid Detection (GARD). The assay is based on chemical stimulation of MUTZ-3 cultures, using the compounds to be assayed as stimulatory agents. The readout of the assay is a transcriptional quantification of the genomic predictors, collectively termed the GARD Prediction Signature (GPS), using a complete genome expression array. Compounds are predicted as either sensitizers or nonsensitizers by a Support Vector Machine model. In this report, we provide a proof of concept for the functionality of the GARD assay by describing the classification of 26 blinded and 11 nonblinded chemicals as sensitizers or nonsensitizers. Based on these classifications, the accuracy, sensitivity, and specificity of the assay were estimated to 89, 89, and 88%, respectively.

Key words: GARD; *in vitro* assay; predictive assay; chemical sensitizers; skin sensitization; allergic contact dermatitis.

Chemical sensitizers are substances of low molecular weight able to induce sensitization in exposed individuals. The most frequent clinical manifestation of chemical sensitization is allergic contact dermatitis (ACD), which is a type IV cell-mediated hypersensitivity reaction (Kimber *et al.*, 2002). The immunological reaction is initiated by dendritic cells (DCs) in peripheral tissue and has been extensively reviewed (Kaplan *et al.*, 2012; Martin *et al.*, 2011; Vocanson *et al.*, 2009). Key events include

the activation, maturation, and migration of DCs, which eventually induce a T-cell mediated immunological response.

The incidences of hypersensitivity toward chemicals are increasing (Lunder and Kansky, 2000; Nguyen *et al.*, 2008), and it is estimated that ~20% of the European population suffers from ACD (Peiser *et al.*, 2012). Furthermore, ACD acquired from occupational environments has become the first ranking occupational disease in many western countries (Diepgen, 2003). However, chemical sensitizers are not exclusively encountered within certain occupational groups, as the general population is exposed to sensitizers in the form of health and beauty care products, cosmetics, and fragrances. Acquired ACD does in many cases imply a reduction of life quality, and the economic burden for society is considered very high (Diepgen and Coenraads, 1999). Therefore, chemical risk assessment and prediction of substances able to induce sensitization is of great importance.

Historically, prediction of chemical sensitizers has been performed using *in vivo* methods, primarily the Guinea Pig Maximization Assay (GPMT) (Magnusson and Kligman, 1969) and the Local Lymph Node Assay (LLNA) (Basketter *et al.*, 2002). However, the implementation of the seventh Amendment to the Cosmetics Directive (76/768/EEC) imposed a ban on all animal experimentation in the cosmetics industry. In addition, the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) legislation requires that all chemical compounds produced in sufficient annual amounts, i.e., ~30,000 chemicals, are to be tested for their abilities to induce sensitization, thus, constituting a capacity problem for *in vivo* methods for risk assessments. Thus, the need for alternative, animal-free methods for assessment of chemical sensitizers is urgent.

In order to meet this demand, several alternative methods for assessment of sensitizers have been proposed. In particular, the Direct Peptide Reactivity Assay (DPRA) (Gerberick *et al.*, 2004, 2007, 2009), Keratinosens (Natsch, 2010; Natsch *et al.*, 2011) and the Human Cell Line Activation Test (h-CLAT) (Ashikaga *et al.*, 2006; Nukada *et al.*, 2011) has shown initially promising results and have been validated (DPRA) or are currently being evaluated in a formal validation process con-

trolled by the European Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM). Furthermore, VITASENS (Hooyberghs *et al.*, 2008) remains an attractive alternative and is considered for ECVAM validation. However, all of these assays are narrow in their readout, and provide little or overlapping mechanistic information. In addition, the predictive performance of these assays might be considered insufficient for stand-alone tests. Thus, accurate tests assessing complementary mechanistic events are still needed.

Recently, we presented a novel assay for assessment of chemical sensitizers, called Genomic Allergen Rapid Detection (GARD) (Johansson *et al.*, 2011, 2013). GARD is based on the myeloid cell line MUTZ-3 (Hu *et al.*, 1996), acting as an *in vitro* model of DCs (Larsson *et al.*, 2006; Masterson *et al.*, 2002). Following cellular stimulation, chemicals are classified as sensitizers or nonsensitizers by evaluating the transcriptional levels of a set of genomic biomarkers, collectively referred to as the GARD Prediction Signature (GPS). The GPS comprises genes that have been associated with innate immune response signaling, DC maturation, stress responses, and recognition of foreign substances, thus providing a holistic readout and the ability to monitor various key cellular events leading to sensitization.

Initial cross-validation exercises of GARD suggested a high predictive performance. However, no validation of external test data has yet been published. In this report, we describe an in-house validation of GARD, using 26 blinded and 11 nonblinded chemicals, most of which have not been used for assay development. We demonstrate the functionality of GARD and report metrics of predictive performance and reproducibility.

MATERIALS AND METHODS

Chemicals and chemical handling. This report describes the classification of 11 nonblinded compounds and 26 blinded compounds. These compounds will in this report be regarded as three separate test sets, referred to as test sets 1–3. The nonblinded samples of test set 1 were chosen to include a wide range of sensitizing potencies. In addition, they were chosen so as to not include chemicals used during assay development (Johansson *et al.*, 2011). The blinded samples of test sets 2 and 3 were provided by two different third parties. The true identities of blinded samples were not revealed until after data presentation to the respective sample provider. The identities of the samples of test set 2 are disclosed, whereas the test set 3 will remain partially coded throughout this report. Chemicals of test set 1 were purchased from Sigma Aldrich (St Louis, MO). All chemicals were stored and handled according to recommendations provided by the supplier. The 11 nonblinded compounds and the 26 blinded compounds are listed in Table 1, together with solubility and toxicity data. Nonblinded samples were assayed in triplicates, whereas blinded samples were assayed in duplicates.

Sample preparation and data acquisition. All methods regarding the generation of data, including cell maintenance, chemical stimulation of cells, cytotoxicity screening, isolation, and preparation of genetic material and microarray handling were performed according to published protocols (Johansson *et al.*, 2013).

Statistical analysis. All new microarray data was merged with historical training data (Johansson *et al.*, 2011) prior to Robust Multi-array Average (RMA) normalization. Visualization of data was performed with Principal Component Analysis (PCA), constructed in Qlucore Omics Explorer 2.3 (Qlucore AB, Lund, Sweden). A Support Vector Machine (SVM) (Cortes and Vapnik, 1995) was trained on the training data, using the GPS as variable input. SVMs were constructed in R (R Core Team, 2013) with the additional package *e1071* (Dimitriadou *et al.*, 2011), using a linear kernel and default settings. The samples that are being assayed are then evaluated by the trained and frozen SVM, as a test set. Predictive performance on replicate level was evaluated by the area under the Receiver Operating Characteristic (ROC AUC), calculated and visualized in R with the additional package *ROCR* (Sing *et al.*, 2005). The classification of a compound as a sensitizer or a nonsensitizer is based on the SVM prediction value. In this report, two different classification criteria were applied. Firstly, the original GARD uses the definition “If any one replicate of any compound stimulation has an SVM decision value >0 , the compound should be classified as a sensitizer.” This classification criteria was only applied to predict test set 1. Following analysis of nonblinded samples, it was revealed that the predictive accuracy would benefit from a calibration of the SVM output. Thus, using test set 1 as internal calibration controls, all blinded samples were classified using a modified classification criteria defined as “If any one replicate of any compound stimulation has an SVM decision value >-1 , the compound should be classified as a sensitizer.” Microarray data of the training set have been deposited in the Array Express database (<http://www.ebi.ac.uk/arrayexpress/>) with accession number E-MTAB-670.

RESULTS

Visualization of Data

The informational content of the GPS was visualized using PCA, as shown in Figure 1. The PCA is based on the 137 samples used for predictor identification (Johansson *et al.*, 2011), i.e., the training dataset, using the 200 genes of the GPS as variable input. A clear separation between sensitizers and nonsensitizers is observed along the first principal component, which contains 37% of the information distributed over the 200 genes, as previously described. Seemingly, sensitizers and nonsensitizers are separated by a plane along the second and third principal components. Plotting the test sets into this space, without influencing the principal components, it is clear that the separation

TABLE 1
Chemicals Used for In-House Validation of GARD

Sample ID	CAS	LLNA potency	Max solubility ^a (μM)	Rv90 ^a (μM)	GARD input concentration ^a (uM)
<i>Test set 1</i>					
Diphenylcyclopropenone	886-38-4	Extreme	100	5	5
4-Nitrobenzylbromide	100-11-8	Extreme	100	1.5	1.5
4-Methylaminophenol sulfate (Metol)	55-55-0	Strong	200	7	7
Hydroquinone	123-31-9	Strong	500	20	20
Phenylacetaldehyde	122-78-1	Moderate	500	50	50
Diethyl maleate	141-05-9	Moderate	500	120	120
Phenyl benzoate	93-99-2	Weak	200	—	200
Dimethyl isophthalate	1459-93-4	Nonsensitizer	250	—	250
Xylene	1330-20-7	Nonsensitizer	—	—	500
Hexane	110-54-3	Nonsensitizer	—	—	500
Chlorobenzene	108-90-7	Nonsensitizer	98	—	98
<i>Test set 2</i>					
4-Nitrobenzylbromide	100-11-8	Extreme	50	2	2
Methylidibromoglutaronitrile	35691-65-7	Strong	400	20	20
Lauryl gallate	1166-52-5	Strong	50	3	3
2-Mercaptobenzothiazole ^b	149-30-4	Moderate	—	300	300
Cinnamal	104-55-2	Moderate	—	50	50
Tetramethyl thiuram disulfide	137-26-8	Moderate	100	0.1	0.1
Phenyl benzoate	93-99-2	Weak	—	—	500
Sodium lauryl sulfate ^{b,c}	151-21-3	Weak	—	—	500
Salicylic acid ^b	69-72-7	Nonsensitizer	—	—	500
Lactic acid ^b	50-21-5	Nonsensitizer	—	—	500
<i>Test set 3</i>					
Diphenylcyclopropenone	886-38-4	Extreme	100	8	8
Dinitrochlorobenzene ^b	14551-14-0	Extreme	100	3	3
CAPE	104594-70-9	Extreme	100	5	5
Taxifolin	480-18-2	Strong	—	320	320
PPG-6-decyltetradeceth-30	72484-69-6	Moderate	—	300	300
PHMG-HCl	57028-96-3	Moderate	—	0.2	0.2
<i>Coded-2</i>	—	Moderate	—	300	300
<i>Coded-1</i>	—	Weak	—	—	500
7-Hydroxycitronellal	107-75-5	Weak	—	—	500
Pinanediol	22422-34-0	Weak	—	—	500
Benzalkoniumchloride ^c	63449-41-2	Nonsensitizer	100	3	3
4-Chloroaniline ^c	106-47-8	Nonsensitizer	—	—	500
Alpha-glucosylrutin	130603-71-3	Nonsensitizer	—	—	500
Menthol	89-78-1	Nonsensitizer	—	—	500
<i>Coded-3</i>	—	Nonsensitizer	—	2	2
Dimethyl CAPE	14551-14-0	Nonsensitizer	10	10	10

^aThe GARD input concentration is derived from both cell toxicity (Rv90, chemical concentration inducing 90% relative viability) and chemical solubility. For details on methodology, see Johansson *et al.* (2013).

^bIndicated compounds were included in assay development, see Johansson *et al.* (2011).

^cIndicated compounds are false classifications in LLNA. Sodium lauryl sulfate is a false positive (Gerberick *et al.*, 2005). Benzalkoniumchloride and 4-chloroaniline are false negatives (Haneke *et al.*, 2001).

of sensitizers and nonsensitizers follow the same direction as in the training data, i.e., along the first principal component, when observing the nonblinded stimulations of test set 1 (Figs. 1A and 1B). However, it is evident at this point that there is a slight displacement of the samples in the test set. Using the plane along the second and third principal components as a visual classifier would yield a number of false negatives.

Similarly, the blinded test sets 2 and 3 were plotted into the PCA space based on the training dataset (Figs. 1C–F). Attempting to perform visual predictions from Figure 1D, using the plane created by the second and third principal components as a classifier, it is clear that 4-nitrobenzylbromide, 2-mercaptobenzothiazole, cinnamal, and tetramethyl thiuram disulfide cluster with sensitizers of the training dataset. Similarly, salicylic acid, lactic acid, sodium lauryl sulfate, and

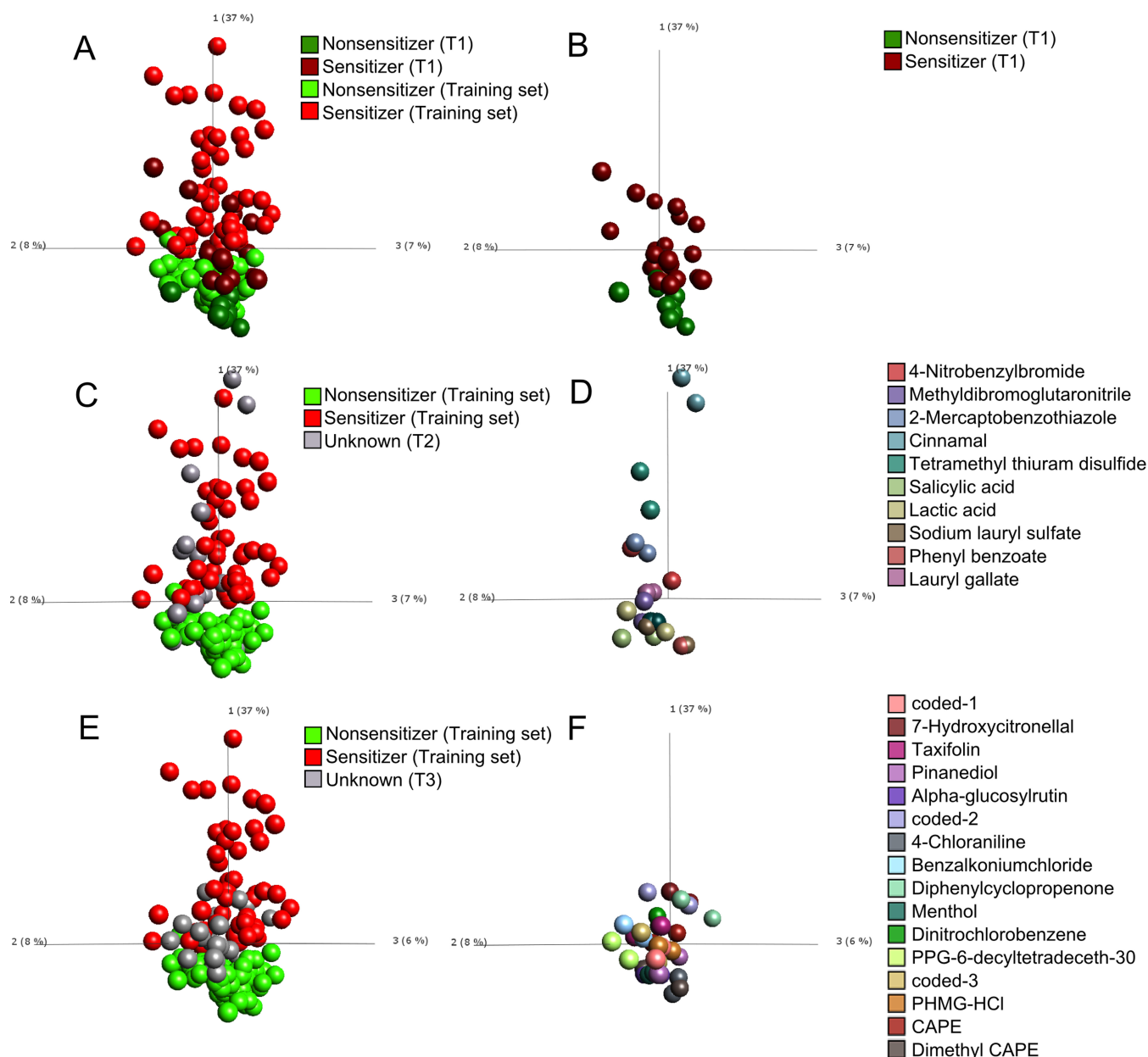


FIG. 1. PCA visualization of microarray data. The three-dimensional PCA space is built up of the first three principal components of the 137 samples used for assay development, using the 200 genes of the GARD Prediction Signature as variable input. The separation of sensitizers (bright red) and nonsensitizers (bright green) can be observed along the first principal component, which contains 37% of the information distributed over the 200 input variables. The samples of the test sets, containing triplicates (test set 1, n stimulations = 11, A and B) or duplicates (test set 2, n stimulations = 10, C and D, test set 3, n stimulations = 16, E and F) of chemical stimulations with known (test set 1, A and B) or unknown (test sets 2 and 3, C–F) sensitizing properties, were plotted into the PCA space without influencing the principal components. In (B), (D), and (F), the training dataset is removed for a clearer view of the separation of samples in the test datasets.

phenyl benzoate are grouped with nonsensitizers of the training data. However, methylidibromoglutaronitrile and lauryl gallate are borderline cases, with samples positioned closely to the plane created by the second and third principal components, impeding visual classifications. Similarly, samples from test set 3 are more densely positioned closed to the classifier plane, although 7-hydroxycitronellal, taxifolin, *coded-2*, diphenylcyclo-

propanone, dinitrochlorobenzene, and CAPE are visually identifiable as sensitizers, based on their position related to sensitizers of the training data, whereas alpha-glucosylrutin, 4-chloroaniline, menthol, and dimethyl CAPE are seemingly nonsensitizers (Fig. 1F).

Support Vector Machine Classifications and Classifier Performance

An SVM was trained on the training data obtained during assay development (Johansson *et al.*, 2011) and applied to predict each sample of the test datasets. Decision values from each prediction are presented in Table 2 and classifier performance is presented in Figure 2. As stated in previously published protocols (Johansson *et al.*, 2013), a compound is classified as a sensitizer if any one replicate stimulation has a decision value >0 . However, as observed in Figures 1A and 1B, a shift in PCA space is apparent when comparing samples of the nonblinded test set with training data. This shift is confirmed in Figure 2B, which shows a local maximum in accuracy at a classifier cutoff of -1 . Importantly, at this point in analysis, only the accuracy of test set 1 was known. Thus, prior to classification of blinded samples, we modified the classifier to classify compounds as sensitizers if any one replicate stimulation has a decision value >-1 . Results from SVM classifications are presented in Table 2. For test set 1, results from both classification criteria are disclosed. For all blinded samples, only the modified classification criteria was used.

As the adjusted classification model performed superior to the original model in the nonblinded test dataset (test set 1), the final predictions of the blinded test dataset samples, as presented to third party sample providers, were based on this model as well. After disclosure of the true identities of tested compounds, predictor performance was calculated. Predictor performance on replicate level are shown in Figure 2A, with ROC AUCs of 0.90, 0.90, and 0.89 for test sets 1, 2, and 3, respectively. Following the classifications as defined by the modified criteria, i.e., after test set 1 was used for SVM calibration, GARD was able to predict the 26 blinded compounds with accuracy, sensitivity, and specificity of 89, 99, and 88%, respectively. Additional Cooper statistics (Cooper *et al.*, 1979) for the separate test datasets are presented in Table 3. These performance metrics are calculated by comparison to LLNA data, except for compounds which are known to be yield false classifications in the LLNA. Such compounds are indicated in Table 1. Of note, the LLNA accuracy calculated using the same dataset is identical (89%), although the misclassifications differ slightly.

Reproducibility

The reproducibility of GARD was assessed by calculating the coefficient of variation (CV) of the SVM prediction values, presented in Table 2. The average CV was 113%, whereas the median CV was 36%. As indicated by the difference in average and median values, a few outliers heighten the average CV. These outliers are marked in red in Table 2. Omitting these outliers, the average CV was 29%.

DISCUSSION

There is currently an urgent need to develop and validate animal-free testing methods for assessment of chemical sensitizers. Numerous assays have been proposed, although few of these methods have yet completed an ECVAM validation process.

Currently, it is anticipated that animal testing will not be replaced with any one standalone *in vitro* assay. Rather, Integrated Testing Strategies (ITS), comprised of a battery of alternative assays is becoming an increasingly appealing alternative (Kinsner-Ovaskainen *et al.*, 2009, 2012). Ideally, an ITS should be based on assays targeting a wide spectra of mechanistic key events of the sensitization process, allowing for educated risk assessments based on weight of evidence. However, accurate decision making will be entirely dependent on the predictive performance of the ITS components. Thus, transparency and publications regarding predictive performance of assays in development are of great importance, as the weight of evidence assigned to each assay in an ITS should be correlated to historical performance.

The GARD is a novel assay based on a genomic biomarker signature, evaluated poststimulation of MUTZ-3 cells. In this report, we describe an in-house validation study, illustrating the functionality of GARD as well as providing metrics of predictive performance. Eleven chemicals were consciously chosen to represent various chemistries and sensitizing potencies for an initial prevalidation of nonblinded samples. With a modification of the GARD classifier, these chemicals were correctly classified with an accuracy of 100%. It should be noted that this figure is not suitable for predictor performance representation, due to risk of overfitting. However, subsequent analysis of two blinded test sets provided by third parties, consisting of 10 and 16 chemicals, revealed accuracies of 90 and 88%, respectively, using the same classification criteria. Thus, an overall accuracy of 89% was observed in blinded samples, with a sensitivity and specificity of 89 and 88%, respectively. Note that for a number of the tested compounds, human data are in many instances lacking. For this reason, metrics on predictive performance have been calculated by comparison with LLNA data, except for compounds where LLNA are known to yield false classifications. Such cases are indicated in Table 1.

Reports of historical performance of the LLNA differ between chemical cohorts, however, an extensive comparison of *in vivo* prediction data state that compared with human clinical data, the accuracy of LLNA was estimated to 72% (Haneke *et al.*, 2001). Similar data for *in vitro* assays in development was recently published in an extensive study, in which DPRA, Keratinosens and MUSST were challenged with 145 chemicals. The accuracy of the three assays ranged between 71 and 80% (Natsch *et al.*, 2013). Using the same data for training of a Bayesian ITS model, the overall accuracy was elevated to 95% when using 21 chemicals as a left-out test set, thereby provid-

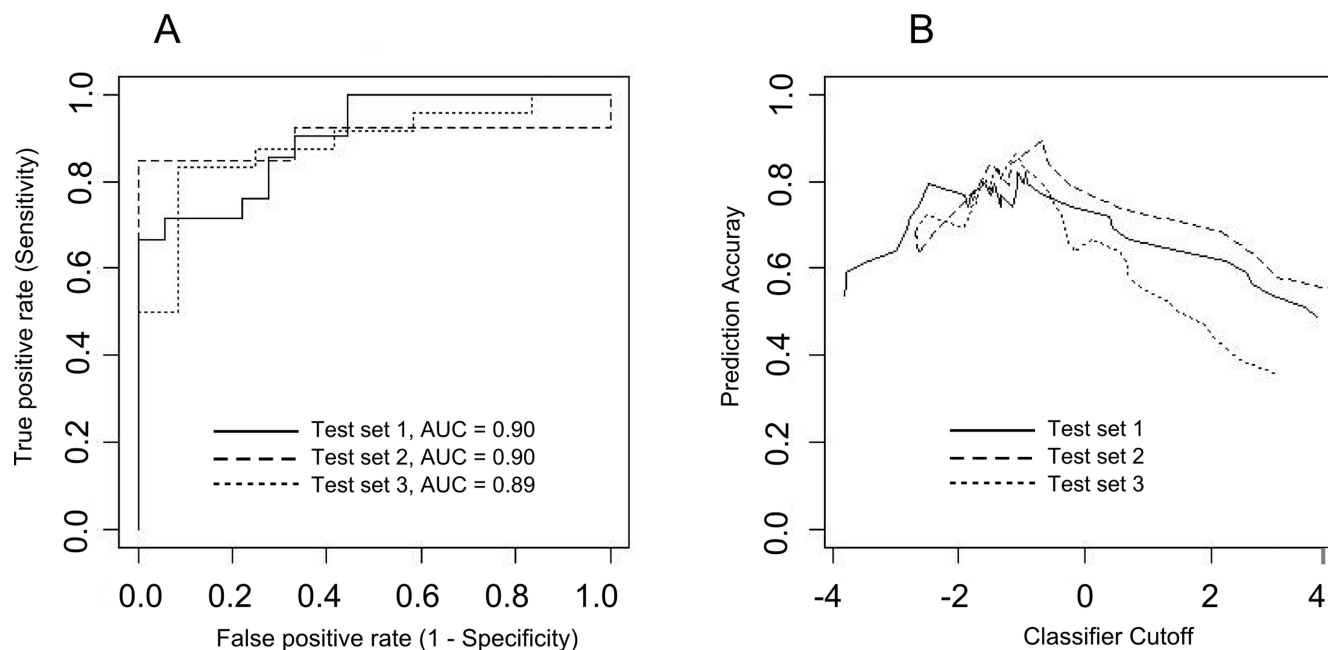


FIG. 2. SVM prediction performance on replicate level. An SVM trained on the training dataset was applied to predict each replicate sample of test sets 1, 2, and 3. (A) Prediction performance evaluated by ROC curves. (B) Prediction accuracy is plotted as a function of SVM decision value cutoff used for classification. ROC, receiver operation characteristic; AUC, area under the curve.

ing a proof of concept of how an ITS can be applied to increase predictive performance (Jaworska *et al.*, 2013).

In the light of these data, the predictive performance of GARD should be considered high. Furthermore, GARD should be considered a highly future-oriented assay in two aspects. Firstly, the use of biomarker signatures makes GARD a flexible assay, meaning that it can evolve and adapt as more knowledge is gained and the understanding of sensitization processes *in vivo* increases. Indeed, as more samples are run, new models can be created, reflecting the increased diversity of reference compounds used for model training. Secondly, the high informational content provided by a holistic genomic readout may be applied for mechanism of action-based investigations, thus enabling the study of specific signaling or metabolic pathways, information which is requested in modern toxicology. Such investigations have been performed using available data, with results indicating a clear correlation of preferential dysregulation of pathways and sensitizer potency (Albrekt *et al.*, 2014). Still, technical issues regarding reproducibility persist, which is addressed below.

During analysis of nonblinded samples, the SVM classifier cutoff was found to benefit from an adjustment. This adjustment of the classifier was needed in order to compensate for a slight misplacement of the testing data in transcriptional space, compared with the training data. Of important note is that no adjustments have been made to the *a priori* defined biomarker signature or the prediction model; only the interpretation of the output was modified. By observing only samples in the test set

in a PCA-space defined by predefined input variables, a separation between sensitizers and nonsensitizers is evident (Fig. 1B). This strengthens our previous claim that the chosen predictors of the GPS indeed decipher information relevant to the biology of chemical skin sensitization.

The apparent shifts in transcriptional levels are most likely due to biological and/or technical variations. The source and frequency of such variations and the impact it might have on routine use of GARD will be the subject of further investigation. Furthermore, whether variation-induced shifts like the one described will always be in the same direction is currently not known. For future experiments, it should be considered essential to apply the use of benchmarked controls, i.e., positive, negative, and intermediate reference compounds with known expected outcomes, similar to the use of test set 1 in this study. In practice, the application of such reference compounds should be considered as a calibration of the output scale, similar to standard curves of single biomarker assays. Similar exercises for calibration of SVMs have previously been described (Jiang *et al.*, 2012). Doing so would allow for defining a decision value range for each dataset, which would enable the possibility to use predefined ratios of the decision value range as classifier cutoffs, rather than using set values.

Having established the relevance of the chosen predictors, it could be argued that the issue could be solved also by a technical platform transfer. Although expression microarrays have been valuable during assay development, allowing for the identification of the genomic predictors of the GPS, the analytical

TABLE 2
Results from SVM Classifications of Test Datasets

Sample ID	SVM decision value			CV ^a (%)	Predicted class ^b	
					pos if one sample	pos if one sample
	1	2	3		>0	>−1
<i>Testset 1</i>						
Diphenylcyclopropanone	1.56	−1.74	0.48	1682	Sensitizer	Sensitizer
4-Nitrobenzylbromide	−1.48	−0.69	−2.61	61	Nonsensitizer	Sensitizer
4-Methylaminophenol sulfate	−1.66	−2.02	−0.96	35	Nonsensitizer	Sensitizer
Hydroquinone	3.44	2.42	3.41	19	Sensitizer	Sensitizer
Phenylacetaldehyde	−0.32	0.52	0.66	185	Sensitizer	Sensitizer
Diethyl maleate	2.85	2.6	2.2	13	Sensitizer	Sensitizer
Phenyl benzoate	−0.82	−0.96	−1.71	41	Nonsensitizer	Sensitizer
Dimethyl Isophthalate	−3.7	−2.92	−3.38	12	Nonsensitizer	Nonsensitizer
Xylene	−1.14	−2.91	−1.68	47	Nonsensitizer	Nonsensitizer
Hexane	−1.26	−1.79	−2.87	42	Nonsensitizer	Nonsensitizer
Chlorobenzene	−3.99	−2.8	−1.53	44	Nonsensitizer	Nonsensitizer
<i>Testset 2</i>						
4-Nitrobenzylbromide	0.78	2.19	—	67	—	Sensitizer
Methyldibromoglutaronitrile	−0.70	−1.31	—	43	—	Sensitizer
Lauryl gallate	−0.38	−0.38	—	0	—	Sensitizer
2-Mercaptobenzothiazole	3.00	3.13	—	3	—	Sensitizer
Cinnamal	11.78	10.53	—	8	—	Sensitizer
Tetramethyl thiuram disulfide	7.71	4.88	—	32	—	Sensitizer
Phenyl benzoate	−2.41	—	—	—	—	Nonsensitizer
Sodium lauryl sulfate	−2.53	−1.21	—	50	—	Nonsensitizer
Salicylic acid	−1.94	−2.52	—	18	—	Nonsensitizer
Lactic acid	−1.42	−1.72	—	14	—	Nonsensitizer
<i>Testset 3</i>						
Diphenylcyclopropanone	1.91	2.29	—	13	—	Sensitizer
Dinitrochlorobenzene	0.78	1.32	—	36	—	Sensitizer
CAPE	−0.40	0.84	—	399	—	Sensitizer
Taxifolin	−0.29	0.44	—	688	—	Sensitizer
PPG-6-decyltetradeceth-30	−0.90	−1.52	—	36	—	Sensitizer
PHMG-HCl	−0.36	−0.45	—	16	—	Sensitizer
Coded-2	2.50	3.07	—	14	—	Sensitizer
Coded-1	−0.86	−1.09	—	17	—	Sensitizer
7-Hydroxycitronellal	2.12	1.64	—	18	—	Sensitizer
Pinanediol	−1.76	−0.43	—	86	—	Sensitizer
Benzalkoniumchloride	0.11	0.65	—	100	—	Sensitizer
4-Chloroaniline	−2.34	−1.23	—	44	—	Nonsensitizer
Alpha-glucosylrutin	−1.36	−2.69	—	46	—	Nonsensitizer
Menthol	−1.77	−1.93	—	6	—	Nonsensitizer
Coded-3	−0.05	−1.24	—	130	—	Sensitizer
Dimethyl CAPE	−1.73	−1.74	—	0	—	Nonsensitizer

^aA CV (coefficient of variation) marked in red indicates divergent measurements between replicates.

^bGreen indicates a correct classification. Red indicates an incorrect classification. Concordance is evaluated compared with LLNA results, except for compounds where LLNA are known to yield false classifications. Such compounds are indicated in Table 1, with supporting references.

performance of microarrays and issues regarding reproducibility between experiments have been under intense debate, with numerous reports of similar cases (Draghici *et al.*, 2006; Liang, 2007; Murphy, 2002). Ideally, a technical platform for routine use of GARD should possess acknowledged low variance and a stable baseline. An investigation of suitable technical platform alternatives for GARD has been initiated. Results show that historical microarray data are successfully reproduced in three sep-

arate high-throughput technical platforms, thus constituting attractive alternatives for routine screenings (Forrery *et al.*, unpublished observations).

In conclusion, we present here a proof of concept study, demonstrating the functionality of GARD as a predictive assay for chemical skin sensitizers. In a chemical cohort consisting of 26 blinded compounds, the accuracy of GARD was estimated to be 89%.

TABLE 3
Cooper Statistics

	Test set 1	Test set 2	Test set 3	Blinded samples	Combined data
Number of correctly classified sensitizers	7/7	6/7	10/11	16/18	23/25
Sensitivity (%)	100	86	91	89	92
Number of correctly classified nonsensitizers	4/4	3/3	4/5	7/8	11/12
Specificity (%)	100	100	80	88	92
Number of correctly classified compounds	11/11	9/10	14/16	23/26	34/37
Accuracy (%)	100	90	88	89	92

FUNDING

Swedish Fund for Research Without Animal Experiments (37/12); Swedish Research Council (2012-2457); Swedish Research Council Formas (2011-1418); Faculty of Engineering (LTH).

ACKNOWLEDGMENT

We would like to thank Ann-Charlott Olsson for microarray sample preparations.

REFERENCES

- 76/768/EEC. http://ec.europa.eu/consumers/sectors/cosmetics/documents/directive/index_en.htm. Accessed March 19, 2014.
- Albrekt, A. S., Johansson, H., Borje, A., Borrebaeck, C., and Lindstedt, M. (2014). Skin sensitizers differentially regulate signaling pathways in MUTZ-3 cells in relation to their individual potency. *BMC Pharmacol. Toxicol.* **15**, 5.
- Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H., and Toyoda, H. (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.
- Basketter, D. A., Evans, P., Fielder, R. J., Gerberick, G. F., Dearman, R. J., and Kimber, I. (2002). Local lymph node assay—Validation, conduct and use in practice. *Food Chem. Toxicol.* **40**, 593–598.
- Cooper, J. A., 2nd, Saracci, R., and Cole, P. (1979). Describing the validity of carcinogen screening tests. *Br. J. Cancer* **39**, 87–89.
- Cortes, C., and Vapnik, V. (1995). Support-Vector Networks. *Mach. Learn.* **20**, 273–297.
- Diepgen, T. L. (2003). Occupational skin-disease data in Europe. *Int. Arch. Occup. Environ. Health* **76**, 331–338.
- Diepgen, T. L., and Coenraads, P. J. (1999). The epidemiology of occupational contact dermatitis. *Int. Arch. Occup. Environ. Health* **72**, 496–506.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., and Weingessel, A. (2011). *e1071: Misc Functions of the Department of Statistics (q1071)*. R package version 1.6. <http://CRAN.R-project.org/package=e1071>. Accessed March 19, 2014.
- Draghici, S., Khatri, P., Eklund, A. C., and Szallasi, Z. (2006). Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet.* **22**, 101–109.
- Gerberick, G. F., Ryan, C. A., Kern, P. S., Schlatter, H., Dearman, R. J., Kimber, I., Patlewicz, G. Y., and Basketter, D. A. (2005). Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis* **16**, 157–202.
- Gerberick, G. F., Troutman, J. A., Foertsch, L. M., Vassallo, J. D., Quijano, M., Dobson, R. L., Goebel, C., and Lepoittevin, J. P. (2009). Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. *Toxicol. Sci.* **112**, 164–174.
- Gerberick, G. F., Vassallo, J. D., Bailey, R. E., Chaney, J. G., Morrall, S. W., and Lepoittevin, J. P. (2004). Development of a peptide reactivity assay for screening contact allergens. *Toxicol. Sci.* **81**, 332–343.
- Gerberick, G. F., Vassallo, J. D., Foertsch, L. M., Price, B. B., Chaney, J. G., and Lepoittevin, J. P. (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. *Toxicol. Sci.* **97**, 417–427.
- Haneke, K. E., Tice, R. R., Carson, B. L., Margolin, B. H., and Stokes, W. S. (2001). ICCVAM evaluation of the murine local lymph node assay. Data analyses completed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. *Regul. Toxicol. Pharmacol.* **34**, 274–286.
- Hooyberghs, J., Schoeters, E., Lambrechts, N., Nelissen, I., Witters, H., Schoeters, G., and Van Den Heuvel, R. (2008). A cell-based in vitro alternative to identify skin sensitizers by gene expression. *Toxicol. Appl. Pharmacol.* **231**, 103–111.
- Hu, Z. B., Ma, W., Zaborski, M., Macleod, R., Quentmeier, H., and Drexler, H. G. (1996). Establishment and characterization of two novel cytokine-responsive acute myeloid and monocytic leukemia cell lines, MUTZ-2 and MUTZ-3. *Leukemia* **10**, 1025–1040.
- Jaworska, J., Dancik, Y., Kern, P., Gerberick, F., and Natsch, A. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. *J. Appl. Toxicol.*, doi:10.1002/jat.2869.
- Jiang, X., Menon, A., Wang, S., Kim, J., and Ohno-Machado, L. (2012). Doubly Optimized Calibrated Support Vector Machine (DOC-SVM): An algorithm for joint optimization of discrimination and calibration. *PLoS One* **7**, e48823.
- 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.
- 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.
- Kaplan, D. H., Igyarto, B. Z., and Gaspari, A. A. (2012). Early immune events in the induction of allergic contact dermatitis. *Nat. Rev. Immunol.* **12**, 114–124.
- Kimber, I., Basketter, D. A., Gerberick, G. F., and Dearman, R. J. (2002). Allergic contact dermatitis. *Int. Immunopharmacol.* **2**, 201–211.
- Kinsner-Ovaskainen, A., Akkan, Z., Casati, S., Coecke, S., Corvi, R., Dal Negro, G., De Bruijn, J., De Silva, O., Gribaldo, L., Griesinger, C., et al. (2009). Overcoming barriers to validation of non-animal partial replacement methods/Integrated Testing Strategies: The report of an EPAA-ECVAM workshop. *Altern. Lab. Anim.* **37**, 437–444.

- Kinsner-Ovaskainen, A., Maxwell, G., Kreysa, J., Barroso, J., Adriaens, E., Alepee, N., Berg, N., Bremer, S., Coecke, S., Comenges, J. Z., *et al.* (2012). Report of the EPAA-ECVAM workshop on the validation of Integrated Testing Strategies (ITS). *Altern. Lab. Anim.* **40**, 175–181.
- Larsson, K., Lindstedt, M., and Borrebaeck, C. A. (2006). Functional and transcriptional profiling of MUTZ-3, a myeloid cell line acting as a model for dendritic cells. *Immunology* **117**, 156–166.
- Liang, P. (2007). MAQC papers over the cracks. *Nat. Biotechnol.* **25**, 27–28, author's reply 28–29.
- Lunder, T., and Kansky, A. (2000). Increase in contact allergy to fragrances: Patch-test results 1989–1998. *Contact Dermat.* **43**, 107–109.
- 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.
- Martin, S. F., Esser, P. R., Weber, F. C., Jakob, T., Freudenberg, M. A., Schmidt, M., and Goebeler, M. (2011). Mechanisms of chemical-induced innate immunity in allergic contact dermatitis. *Allergy* **66**, 1152–1163.
- Masterson, A. J., Sombroek, C. C., De Gruijl, T. D., Graus, Y. M., Van Der Vliet, H. J., Loughheed, S. M., Van Den Eertwegh, A. J., Pinedo, H. M., and Scheper, R. J. (2002). MUTZ-3, a human cell line model for the cytokine-induced differentiation of dendritic cells from CD34+ precursors. *Blood* **100**, 701–703.
- Murphy, D. (2002). Gene expression studies using microarrays: Principles, problems, and prospects. *Adv. Physiol. Educ.* **26**, 256–270.
- 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.
- Natsch, A., Caroline, B., Leslie, F., Frank, G., Kimberly, N., Allison, H., Heather, I., Robert, L., Stefan, O., Hendrik, R., *et al.* (2011). The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers in vitro: Results of a ring-study in five laboratories. *Toxicol. In Vitro* **25**, 733–744.
- Nguyen, S. H., Ryan, C. A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., and Kern, P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *J. Appl. Toxicol.* doi: 10.1002/jat.2868.
- Nguyen, S. H., Dang, T. P., Macpherson, C., Maibach, H., and Maibach, H. I. (2008). Prevalence of patch test results from 1970 to 2002 in a multi-centre population in North America (NACDG). *Contact Dermat.* **58**, 101–106.
- Nukada, Y., Ashikaga, T., Sakaguchi, H., Sono, S., Mugita, N., Hirota, M., Miyazawa, M., Ito, Y., Sasa, H., and Nishiyama, N. (2011). Predictive performance for human skin sensitizing potential of the human cell line activation test (h-CLAT). *Contact Dermat.* **65**, 343–353.
- Peiser, M., Tralau, T., Heidler, J., Api, A. M., Arts, J. H., Basketter, D. A., English, J., Diepgen, T. L., Fuhlbrigge, R. C., Gaspari, A. A., *et al.* (2012). Allergic contact dermatitis: Epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. Current knowledge assembled at an international workshop at BfR, Germany. *Cell. Mol. Life Sci.* **69**, 763–781.
- R Core Team (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL <http://www.R-project.org/>. Accessed March 20, 2014.
- Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCr: Visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941.
- Vocanson, M., Hennino, A., Rozieres, A., Poyet, G., and Nicolas, J. F. (2009). Effector and regulatory mechanisms in allergic contact dermatitis. *Allergy* **64**, 1699–1714.