



From genome-wide arrays to tailor-made biomarker readout – Progress towards routine analysis of skin sensitizing chemicals with GARD



Andy Forreryd^a, Kathrin S. Zeller^a, Tim Lindberg^a, Henrik Johansson^b, Malin Lindstedt^{a,*}

^a Department of Immunotechnology, Medicon Village (406), 22381 Lund, Sweden

^b SenzaGen AB, Medicon Village (406), 22381 Lund, Sweden

ARTICLE INFO

Article history:

Received 19 July 2016

Accepted 9 September 2016

Available online 13 September 2016

Keywords:

GARD

In vitro assay

Predictive genomic biomarker signature

Skin sensitization

ABSTRACT

Allergic contact dermatitis (ACD) initiated by chemical sensitizers is an important public health concern. To prevent ACD, it is important to identify chemical allergens to limit the use of such compounds in various products. EU legislations, as well as increased mechanistic knowledge of skin sensitization have promoted development of non-animal based approaches for hazard classification of chemicals. GARD is an *in vitro* testing strategy based on measurements of a genomic biomarker signature. However, current GARD protocols are optimized for identification of predictive biomarker signatures, and not suitable for standardized screening.

This study describes improvements to GARD to progress from biomarker discovery into a reliable and cost-effective assay for routine testing. Gene expression measurements were transferred to NanoString nCounter platform, normalization strategy was adjusted to fit serial arrival of testing substances, and a novel strategy to correct batch variations was presented. When challenging GARD with 29 compounds, sensitivity, specificity and accuracy could be estimated to 94%, 83% and 90%, respectively.

In conclusion, we present a GARD workflow with improved sample capacity, retained predictive performance, and in a format adapted to standardized screening. We propose that GARD is ready to be considered as part of an integrated testing strategy for skin sensitization.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hallmarks of sensitization to small reactive chemical substances include activation of dendritic cells (DCs) and initiation of adaptive immune responses. Upon repeated exposure to the same compound, specific Th1 and cytotoxic CD8+ T-cells elicit an inflammatory reaction in the skin, resulting in development of clinical symptoms, such as allergic contact dermatitis (ACD) (Akhavan & Cohen, 2003; Fonacier, Dreskin, & Leung, 2010). Chemical allergens, currently exceeding 4000 known substances (DeGroot, 2008) are present in a variety of products found in households and at workplaces. Common examples include fragrances, cosmetics (Peiser et al., 2012), and certain vulcanized rubber products, such as disposable rubber gloves (Bergendorff, Persson, Ludtke, & Hansson, 2007; Geier, Lessmann, Uter, & Schnuch, 2003). The incidence of ACD is increasing as a consequence of frequent exposure to these products (Lunder & Kansky, 2000; Nguyen, Dang,

MacPherson, Maibach, & Maibach, 2008). ACD is the most frequent manifestation of immunotoxicity in humans (Kimber, Basketter, Gerberick, & Dearman, 2002), affecting approximately 20% of the general population (Thyssen, Linneberg, Menne, & Johansen, 2007), and is especially problematic within certain occupational groups, such as health care workers and cleaners (Dickel, Kuss, Blesius, Schmidt, & Diepgen, 2001).

To limit exposure to hazardous chemical allergens, the European Union (EU) has enforced legislations such as REACH (EU, 2006), which requires risk assessment of all chemicals produced >1 ton/year. Identification of chemical allergens is routinely performed with animal models, such as the murine local lymph node assay (LLNA) (Basketter et al., 2002). Since 2013, animal testing of cosmetic products, or ingredients thereof is banned within the EU (EU, 2009), and similar initiatives have been enforced, or are currently under the investigation in several countries. Such legislations, together with detailed knowledge of key mechanistic events linked to ACD, summarized in an adverse outcome pathway (AOP) (OECD, 2012), has challenged the traditional use of high-dose animal models, resulting in a paradigm shift within the field of toxicity testing, to promote the use of knowledge- and mode of action-based testing using *in vitro* models.

To this end, all key mechanistic events in the AOP for skin sensitization have been targeted for the development of various *in vitro*, *in*

* Corresponding author at: Lund University, Medicon Village (406), 22381 Lund, Sweden.

E-mail addresses: andy.forreryd@immun.lth.se (A. Forreryd), kathrin.zeller@immun.lth.se (K.S. Zeller), tim.lindberg@immun.lth.se (T. Lindberg), henrik.johansson@senzagen.com (H. Johansson), malin.lindstedt@immun.lth.se (M. Lindstedt).

chemico or *in silico* predictive models, as reviewed in (Reisinger et al., 2015). The direct peptide reactivity assay (DRPA) (Gerberick et al., 2004), KeratinoSens™ (Natsch, 2010), and the human cell line activation test (h-CLAT) (Ashikaga et al., 2006; Nukada et al., 2011), have been validated by European Reference Laboratory for alternatives to animal testing (EURL ECVAM) for use within Integrated Approaches to Testing and Assessment (IATA). These assays are based on single or few biomarkers, providing limited mechanistic information. Thus, there is still a need for reliable assays with higher informational content, targeting complementary mechanistic events.

Genomic allergen rapid detection (GARD) (Johansson, Albrekt, Borrebaeck, & Lindstedt, 2013; Johansson, Lindstedt, Albrekt, & Borrebaeck, 2011) developed by our group is based on measurements of transcriptional levels of 200 genomic biomarkers, collectively referred to as the GARD prediction signature (GPS), in a myeloid cell line using transcriptome-wide microarray profiling. Transcripts in the GPS are involved in cellular pathways associated with antigen recognition, DC maturation and cellular stress, thus immunologically relevant to the sensitization process. Upon cellular stimulation with an unknown compound, the induced changes in gene expression levels within the GPS are compared to profiles of reference chemicals using supervised machine learning based on support vector machines (SVM) (Cortes & Vapnik, 1995). In a recent *in-house* validation study, the accuracy of GARD was estimated to 89% for classification of 26 blinded compounds (Johansson et al., 2014). However, the microarray platform used for identification of transcripts in GPS is not well-adapted for routine testing of a large number of chemicals, due to limitations in sample throughput, high costs and complex protocols. Furthermore, several groups have reported reproducibility issues between experiments performed on the microarray platform (Draghici, Khatri, Eklund, & Szallasi, 2006; Liang, 2007; Murphy, 2002). Thus, we have evaluated the performance of three medium-throughput platforms on a subset of transcripts from the GPS (Forreryd, Johansson, Albrekt, & Lindstedt, 2014), and showed that the NanoString nCounter platform (Geiss et al., 2008) was a promising amplification-free alternative to microarrays. The platform allows for specific detection of individual transcripts in a multiplex format with high sample throughput, while providing a resource-effective, simple and highly automated protocol with little hands-on time.

In this study, we present a novel GARD workflow to meet industrial and regulatory demands for reliability, cost effectiveness and sample capacity. In a first step, we transferred the gene expression measurements from transcriptome-wide profiling using microarrays, into the NanoString nCounter platform, measuring only the transcripts in the GPS biomarker signature. In a second step to further improve robustness and reliability in classifications, we modified the normalization strategy, evaluated a selection of machine learning algorithms, and introduced a novel approach to adjust for batch variations. As a final proof of concept to validate the novel workflow, we performed an *in-house* validation study involving the classification of 29 independent test compounds. Based on results in this study, we here present a reliable and robust experimental design, establishing the applicability of the GARD assay for routine screening of skin sensitizing compounds.

2. Material and methods

2.1. Chemicals and datasets

A total of 38 chemical compounds, including 18 well-defined skin sensitizing compounds and 20 non-sensitizing compounds (Table 1), as well as the vehicle controls water and dimethyl sulfoxide (DMSO) were used for stimulation of cells in order to generate the training set for supervised machine learning and for construction of predictive models. Compounds were identical to those used during identification of the GARD prediction signature, as described in Johansson, Lindstedt, Albrekt, & Borrebaeck (2011). In addition to the training set, a total of

29 chemical compounds were used for stimulation of cells in order to generate independent test sets for validation of predictive models (Table 2). The 29 compounds were further divided into two separate datasets, referred to as test set 1 and test set 2. Test set 1 comprised a total of 20 chemical compounds, including 16 compounds previously classified in GARD (Johansson et al., 2014). Test set 2 comprised a total of nine chemical compounds, all previously unseen by the model, and none of the compounds previously classified by GARD. In addition, each test set included a selection of internal control compounds from the training data set, so-called benchmark control samples. Such compounds included in test set 1 and test set 2 were collectively referred to as benchmark set 1 ($n = 12$), and benchmark set 2 ($n = 10$), respectively (Table 2). All chemical compounds in this study were purchased from Sigma Aldrich (St. Louis, MO, USA), except the chemicals DMTBS and DETBS, which were a kind gift from Dr. Ola Bergendorff, Dept. of Occupational and Environmental Dermatology at Lund University, Sweden.

2.2. Cell handling, chemical stimulations, phenotypic analysis and total RNA isolation

Each chemical compound was dissolved in appropriate vehicle (water or DMSO), based on solubility properties, and used for stimulations of cells at a single concentration, called the GARD input concentration. These concentrations were determined based on cytotoxicity and solubility of individual compounds, using previously established protocols, as described in Johansson, Albrekt, Borrebaeck, & Lindstedt (2013). The myeloid cell line MUTZ-3 was purchased from Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany). All cellular stimulations were performed in triplicate reactions, using a new batch of cells for each individual replicate, generating a total dataset comprising 273 stimulations (training set = 120, test set 1 = 60, test set 2 = 27, benchmark set 1 = 36, benchmark set 2 = 30). Handling of cells, including preparation of cells prior to stimulations, quality control of cells, flow cytometry analysis, and chemical stimulations were performed in accordance with previously published protocols (Johansson, Lindstedt, Albrekt, & Borrebaeck, 2011; Johansson, Albrekt, Borrebaeck, & Lindstedt, 2013). After 24 h of cellular stimulations, cells were harvested, lysed in TRIzol reagent (Thermo Scientific, Waltham, MA), and stored at -20°C until RNA extraction. Total RNA was isolated from TRIzol reagent using Direct-zol™ RNA MiniPrep column purification kit (Zymo research, Irvine, CA) according to protocols provided by manufacturer, and concentrations were determined using Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Total RNA was stored at -80°C until analysis.

2.3. CodeSet design: selection of probes for endogenous and candidate reference transcripts

NanoString nCounter probe pairs (NanoString Technologies, Seattle, WA) targeting endogenous transcripts in the GPS were carefully designed to match at exon level to the transcripts of the Affymetrix HuGene 1.0 ST array measurements (Johansson, Lindstedt, Albrekt, & Borrebaeck, 2011). Each probe pair was translated to the 100 bp target sequence using Affymetrix exon probe set IDs as identifier for a specific target exon. In cases where several different exons were possible targets for probe selection to an individual transcript, design priority was given to exons giving rise to highest signal on the Affymetrix HuGene 1.0 ST array measurements, at the same time falling within the kinetic parameters of the nCounter system, and not cross-reacting with non-target transcripts. In total, 196 transcripts could be matched to appropriate Nanostring probe IDs and included in the custom CodeSet (Supplementary Table 1). Candidate reference transcripts were selected from the nCounter human reference GX kit, or selected based on low expression variance from available Affymetrix HuGene 1.0 ST array measurements

Table 1
Chemical compounds included in the training set.

Compound	Abbreviation	Chemical abstract service (CAS)	Vehicle	GARD input concentration (μM)
Skin sensitizers				
2,4-Dinitrochlorobenzene	DNCB	97-00-7	DMSO	4
2-Aminophenol	2AP	95-55-6	DMSO	100
2-Hydroxyethyl acrylate	2HA	818-61-1	ddH ₂ O	100
2-Mercaptobenzothiazole	MBT	149-30-4	DMSO	250
2-Nitro-1,4-phenylenediamine	NPDA	5307-14-2	DMSO	300
α-Hexylcinnamic aldehyde	HCA	101-86-0	DMSO	32
Cinnamyl alcohol	CALD	104-54-1	DMSO	500
Ethylenediamine	EDA	107-15-3	ddH ₂ O	500
Eugenol	EU	97-53-0	DMSO	300
Formaldehyde solution	FA	50-00-0	ddH ₂ O	80
Geraniol	GER	106-24-1	DMSO	500
Glyoxal solution	GO	107-22-2	ddH ₂ O	300
Isoeugenol	IEU	97-54-1	DMSO	300
Kathon CG/ICP	KCD	– ^a	ddH ₂ O	0.0035%
Penicillin G	PEN G	113-98-4	ddH ₂ O	500
Potassium dichromate	PD	7778-50-9	ddH ₂ O	1.5
p-Phenylenediamine	PPD	106-50-3	DMSO	75
Resorcinol	RC	108-46-3	ddH ₂ O	500
Non-sensitizers				
1-Butanol	BUT	71-36-3	DMSO	500
4-Aminobenzoic acid	PABA	150-13-0	DMSO	500
4-Hydroxybenzoic acid	HBA	99-96-7	DMSO	250
Benzaldehyde	BA	100-52-7	DMSO	250
Chlorobenzene	CB	108-90-7	DMSO	98
Diethyl phthalate	DP	84-66-2	DMSO	50
Ethyl vanillin	EV	121-32-4	DMSO	500
Glycerol	GLY	56-81-5	ddH ₂ O	500
Isopropanol	IP	67-63-0	ddH ₂ O	500
Lactic acid	LA	50-21-5	ddH ₂ O	500
Methyl salicylate	MS	119-36-8	DMSO	500
N,N-Dimethylformamide	DMF	68-12-2	ddH ₂ O	500
Octanoic acid	OA	124-07-2	DMSO	500
Phenol	PHE	108-95-2	ddH ₂ O	500
Potassium permanganate	PP	7722-64-7	DMSO	38
Propylene glycol	PG	57-55-6	ddH ₂ O	500
Salicylic acid	SA	69-72-7	DMSO	500
Sodium dodecyl sulphate	SDS	151-21-3	ddH ₂ O	200
Tween 80	T80	9005-65-6	DMSO	500
Zinc sulphate solution	ZS	7733-02-0	ddH ₂ O	126

^a Kathon CG is a mixture of the active ingredients 5-chloro-2-methyl-4-isothiazolin-3-one (CAS: 26172-55-4) and 2-methyl-4-isothiazolin-3-one (CAS: 2682-20-4). Kathon is a registered trademark of Rohm and Hass (Philadelphia, PA, USA).

(Supplementary Table 2). The CodeSet, comprising custom designed endogenous probe pairs and candidate reference probe pairs, and master kits containing all reagents used to perform gene expression measurements on nCounter, were obtained from NanoString Technologies.

2.4. Gene expression measurements on NanoString nCounter

Gene expression measurements were performed on the NanoString GEN2 nCounter Analysis System comprising the GEN2 Prep Station 5s for post-hybridization processing, and the Digital Analyzer 5s for data collection. All steps were performed according to standardized protocols for gene expression measurements as provided by NanoString. CodeSet and hybridization buffer were mixed with 100 ng total RNA (5 μl at 20 ng/μl) and incubated at 65 °C for 24 h. After hybridization, samples were transferred to the Prep Station and processed using the High Sensitivity protocol. At conclusion of protocol, cartridges were sealed with cartridge adhesive cover, transferred to the Digital Analyzer and matched to the Cartridge Definition File (CDF). Data collection was initiated immediately, using maximal data resolution (555 Fields of view). Data from all runs (n = 273) were imported into nSolver version 2.0 (NanoString Technologies), matched to corresponding Reporter Library File (RLF), and subjected to quality control using the software quality metrics with default cutoff criteria.

2.5. Data pre-processing, selection of reference genes and normalization

Two different normalization methods were applied to the nCounter data. In the first strategy, raw gene expression data was normalized according to protocols recommended by NanoString in the nCounter Expression Data Analysis guide. In short, raw data was adjusted for platform associated sources of variation using geometrical average of counts from positive control probes, and non-specific background was adjusted by subtracting average counts for negative control probes (using counts <25 as a threshold for including a specific negative probe) in each sample. The geometric mean of reference genes was used for correction of differences in RNA input between samples. The best performing reference genes were evaluated from the set of candidate reference genes (n = 24, Supplementary Table 2) on samples in training set. The GeNorm (Vandesompele et al., 2002) and NormFinder (Andersen, Jensen, & Orntoft, 2004) algorithms were used for rankings. GeNorm calculations were performed using an in-house modified script implemented into R statistical environment version 3.2.4 (R Development Core Team, 2013), and rankings were based on average stability values (M) of remaining candidate reference genes during stepwise exclusion of the least stable gene. NormFinder calculations were performed using the R script provided by Molecular Diagnostic Laboratory, Dept. of Molecular Medicine (Aarhus University Hospital Skejby, Denmark), and

Table 2

Chemical compounds included in benchmark sets and independent test sets for validation of GPS.

Compound	Abbreviation	Class	CAS	Vehicle	GARD input (μM)
Benchmark set 1 and 2					
2,4-Dinitrochlorobenzene	DNCB	Sensitizer	97-00-7	DMSO	4
2-Aminophenol	2AP	Sensitizer	95-55-6	DMSO	100
2-Hydroxyethyl acrylate	2HA	Sensitizer	818-61-1	ddH ₂ O	100
Geraniol	GER	Sensitizer	106-24-1	DMSO	500
Isoeugenol ^a	IEU	Sensitizer	97-54-1	DMSO	300
p-Phenylenediamine	PPD	Sensitizer	106-50-3	DMSO	75
Resorcinol	RC	Sensitizer	108-46-3	ddH ₂ O	500
1-Butanol	BUT	Non-Sensitizer	71-36-3	DMSO	500
Benzaldehyde ^a	BA	Non-Sensitizer	100-52-7	DMSO	250
Chlorobenzene	CB	Non-Sensitizer	108-90-7	DMSO	98
Dimethyl sulfoxide	DMSO	Vehicle control	67-68-5		
Double-distilled water	ddH ₂ O	Vehicle control			
Testset 1					
2-Nitro-1,4-phenylenediamine	NPDA	Sensitizer	5307-14-2	DMSO	300
4-Nitrobenzyl bromide	4NBB	Sensitizer	100-11-8	DMSO	5
7-Hydroxycitronellal	7HC	Sensitizer	107-75-5	DMSO	500
Benzalkonium chloride ^b	BZK	Sensitizer	63449-41-2	DMSO	3
Diethyl maleate	DEM	Sensitizer	141-05-9	DMSO	200
Diphenylcyclopropenone	DPC	Sensitizer	886-38-4	DMSO	8
Ethyl acrylate	EA	Sensitizer	140-88-5	DMSO	500
Hydroquinone	HQ	Sensitizer	123-31-9	DMSO	45
Lauryl gallate	LG	Sensitizer	1166-52-5	DMSO	3
Phenylacetaldehyde	PA	Sensitizer	122-78-1	DMSO	75
Tetramethylthiuram disulfide	TMTD	Sensitizer	137-26-8	DMSO	0.1
Dimethyl isophthalate	DMI	Non-Sensitizer	1459-93-4	DMSO	250
Hexane	HEX	Non-Sensitizer	110-54-3	DMSO	500
Lactic acid ^c	LA	Non-Sensitizer	50-21-5	ddH ₂ O	500
Menthol	MEN	Non-Sensitizer	89-78-1	DMSO	500
Methyl salicylate ^c	MS	Non-Sensitizer	119-36-8	DMSO	500
Saccharin	SAC	Non-Sensitizer	81-07-02	DMSO	500
Salicylic acid ^c	SA	Non-Sensitizer	69-72-7	DMSO	500
Sodium dodecyl sulphate ^{c,b}	SDS	Non-Sensitizer	151-21-3	ddH ₂ O	500
Xylenes	XYL	Non-Sensitizer	1330-20-7	DMSO	500
Testset 2					
2,4-Dinitrofluorobenzene	DNFB	Sensitizer	70-34-8	DMSO	10
3-Methylcatechol	3MCAT	Sensitizer	488-17-5	DMSO	40
Butyl glycidyl ether	BGE	Sensitizer	2426-08-6	DMSO	500
Diethylthiocarbamyl benzothiazole sulfide ^d	DETBS	Sensitizer	95-30-7	DMSO	30
Dimethylthiocarbamyl benzothiazole sulfide ^d	DMTBS	Sensitizer	3432-25-5	DMSO	8
Lyril	LYR	Sensitizer	31906-04-4	DMSO	100
1-Bromobutane	1BB	Non-Sensitizer	109-65-9	DMSO	500
Benzoic acid	BEA	Non-Sensitizer	65-85-0	DMSO	500
Citric acid	CIA	Non-Sensitizer	77-92-9	DMSO	500

^a Compounds included only in benchmark control samples for test set 1.^b Compounds are false classifications in LLNA. Benzalkonium chloride is a false negative (Haneke, Tice, Carson, Margolin, & Stokes, 2001), and sodium dodecyl sulphate is a false positive (Gerberick et al., 2005).^c Compounds also present in the training set.^d Compounds have not been classified in LLNA. Classifications are based on positive results in human patch testing (Hansson, Ponten, Svedman, & Bergendorff, 2014).

rankings were based on stability values, taking possible variations between subgroups of samples into considerations by using annotations for sensitizers and non-sensitizers. In the second normalization strategy, raw nCounter gene expression data was normalized using an in-house developed single-sample normalization strategy implemented into R statistical environment, called Counts per Total Counts (CPTC). The method reported normalized values for each gene in a certain sample as the ratio of the count of the specific gene to the total counts of all endogenous genes within the GPS of the same sample.

2.6. SVM classifications and estimate of predictive performances

The most suitable normalization strategy for GARD was evaluated by estimating predictive performances of support vector machine (SVM) models (Cortes & Vapnik, 1995), using a leave-one-out cross-validation (LOOCV) procedure. SVMs were compiled in R statistical environment, using the additional package e1071 (Dimitriadou, Hornik, Leisch, Meyer, & Weingessel, 2011). In short, SVM models, based on linear kernels and default settings, were

repeatedly trained on all but one sample, and the left out sample was set aside to estimate model performance. The process was iterated until all samples had been assigned a SVM decision value (SVM DV). Decision values for all samples were collected and used to generate a receiver operating characteristics (ROC) curve, and the area under the curve (AUC) served as measure of model performance (Lasko, Bhagwat, Zou, & Ohno-Machado, 2005). Calculations and visualizations were performed in R statistical environment, using the additional package ROCR (Sing, Sander, Beerenwinkel, & Lengauer, 2005). The Predictive performance of GPS was estimated using the independent test sets, as described in Johansson et al. (2014). Briefly, the SVM model described above was again trained on samples in the training set, using GPS as variable input, and applied to classify samples in test sets. Prior to final classification of samples in test sets, the SVM model was calibrated on internal control samples in the corresponding benchmark set. During model calibration, the trained SVM was applied to predict sensitizing individual replicate sample a SVM DV. Predictions were compared to true sensitizing properties for each compound, and the accuracy

was evaluated by ROC AUC as a function of SVM DV cutoff used for classification. The trained SVM was then again applied to predict samples in test sets. SVM DVs for each sample in the test sets were then scaled by subtracting the discrimination threshold, which maximized model performance for prediction of samples in the corresponding benchmark set. Final classification of a certain chemical was then based on the median SVM DV value for the triplicate stimulations; hence a specific compound was classified as a sensitizer if the median output value of replicates >0 . Accuracy, sensitivity and selectivity of predictions were determined using cooper statistics (Cooper, Saracci, & Cole, 1979). Principal component analysis (PCA) plots and heatmaps were constructed in Qlucore Omics Explorer 3.2 (Qlucore AB, Lund, Sweden).

2.7. Predictive performances of alternative classification models

The predictive performance of GPS was also estimated using the independent test sets on a selection of supervised machine learning algorithms. Generation of binary predictive models, additional pre-processing of data, model tuning and classifications were performed in R statistical environment, using the caret package (Kuhn et al., 2016). Prior to model construction, gene expression data for predictors within GPS was centered round a mean value of zero, and scaled to a standard deviation of one. Tuning parameters for each model was determined using a resampling procedure of 10-fold cross-validation (repeated 15 times) on samples in the training set prior to prediction of samples in independent test sets. Predicted class probability values were collected and used to generate a ROC curve, and the best performing model from resampling results across tuning parameters was determined based on highest ROC AUC value. For the KNN model, the optimal number of neighbors (k) was determined to 61. The partial least squares (PLS) model was constructed using the additional package pls (Mevik & Hovde Liland, 2015). The optimal number of components ($ncomp$) to be used in the classification model was determined to 3. The random forest (RF) (Breiman, 2001) model was constructed using the additional package randomForest (Liaw & Wiener, 2002). The number of trees ($ntrees$) used in the final model was set to 5000, and the optimal number of variables randomly sampled as candidates at each split ($mtry$) was determined to 3. Additional parameters for each model were kept at default settings. The process of training, calibration and final classification of samples was performed as described for the SVM classifications, using the class probability values generated by each model for classifications. Calibrated class probabilities for individual samples of each chemical were summarized in series of box and whisker plots generated in R statistical environment, with the additional package ggplot2 (Wickham, 2009). For reproducibility of models based on stochastic subsampling, the `set.seed` function within R was used.

3. Results

3.1. Evaluation of appropriate normalization strategies

To investigate the possibility to directly implement nCounter measurements into current GARD protocols, we designed a custom NanoString CodeSet to match targets in the GPS biomarker signature. Following standardized protocols of GARD, reference chemicals in the training set (Table 1) were used for stimulation of cells, and chemically-induced changes on transcriptional levels within the GPS were investigated using the nCounter platform. Prior to data analysis, two different strategies for normalization were evaluated. NanoString currently recommends a normalization strategy, which involves adjustment to a set of stably expressed reference genes. Thus, the performance of a selection of candidate reference genes ($n = 24$, Supplementary Table 2) was first evaluated using geNorm and BestFinder algorithms (Supplementary Figure 1). The five most stable reference genes from NormFinder rankings (*GNB1*, *ABCF1* (3), *ARF1MIR3620*, *SDHA* and

ABCF1 (2)) were selected for normalization of data. As an alternative to the reference normalization strategy, gene expression data was also normalized using our in-house developed single sample normalization method, called Counts per Total Counts (CPTC). In Fig. 1A and B, PCA was used to summarize gene expression data from the different normalization strategies, using GPS as variable input. Expression profiles from reference gene normalized data (Fig. 1A) and CPTC normalized data (Fig. 1B) formed similar patterns, with a partial separation of sensitizers and non-sensitizers along the first PCA component. The formation of separate clusters indicated that nCounter was able to reproduce the differences in expression profiles between sensitizers and non-sensitizers observed during microarray studies on the same reference set of chemicals, regardless of normalization strategy. In Fig. 1C and D, the unsupervised visualizations were supplemented with a supervised evaluation of the predictive capacity of each normalization strategy, using a SVM model in combination with a LOOCV procedure on samples in the training set. As indicated in the figure, the predictive performance of reference gene normalized data (Fig. 1C) and CPTC normalized data (Fig. 1D) were identical. The high ROC AUC values (ROC AUC = 0.92) obtained from each model indicated a potent ability of both strategies to distinguish sensitizers from non-sensitizers, and confirms the visual findings of partly separated groups in the PCA plots. Considering logistic benefits of normalizing samples individually rather than in groups for chemical testing using GARD, the single sample CPTC normalization was used for all downstream analysis.

3.2. Calibrations of SVM classification models for maximal predictive performance

The ability of GPS to accurately classify unknown test compounds in a model based on nCounter measurements was evaluated using test set 1 and test set 2 (Table 2). Compounds in the test sets were predicted blindly and comprised novel stimulations, processed in separate batches from each other and from the training set. Due to previously observed issues when combining different batches of experiments, we introduced the concept of including samples with known sensitizing properties into each test set, so-called benchmark control samples. The benchmark control samples included in test set 1 and 2 were referred to as benchmark set 1 and 2 (Table 2), respectively. In Fig. 2, PCA was used to visually investigate possible shifts in transcriptional levels between the training set and benchmark sets. Samples in the training set were used to construct the PCA space, components were frozen, and samples in benchmark set 1 (Fig. 2A) and benchmark set 2 (Fig. 2B) were plotted into the frozen model. As evident from the figure, separation of sensitizers and non-sensitizers follows the first PCA component for samples in both training and benchmark sets. The unsupervised PCA visualizations were verified using supervised SVM classifications of benchmark samples. Results from SVM predictions are illustrated in Fig. 2C. The cutoff threshold for maximum performance for prediction of samples in benchmark set 1 was determined to 1.01, and for benchmark set 2 to 0.23. Thus, SVM predictions indicate that samples processed in both benchmark sets would require calibration prior to comparison to historical data, in order not to generate false positive classifications.

3.3. SVM classifications and estimate of predictive performance using independent test sets

Prior to supervised classifications of samples in test set 1 and 2, an attempt was made to perform visual classifications. The PCA models used for visualization of benchmark sets (Fig. 2), were applied also for visualization of samples in independent test sets. In Fig. 3, samples in test set 1 (Fig. 3A) and test set 2 (Fig. 3B), at the time blinded to the model, were plotted into the frozen PCA space. Upon disclosure of true identities of samples in test set 1 (Fig. 3C) and 2 (Fig. 3D), a separation along the first PCA component was once again evident between

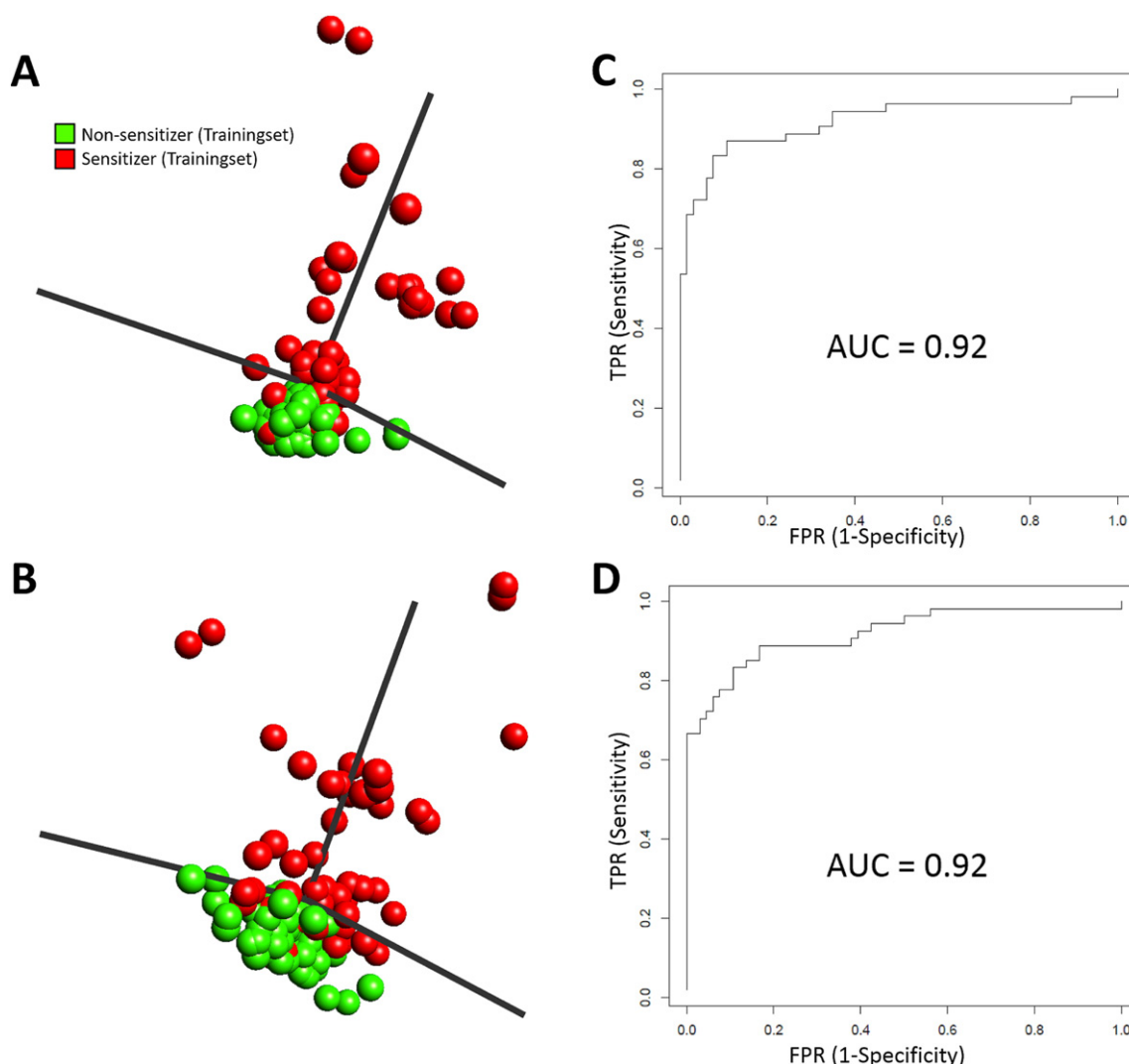


Fig. 1. Validation of predictive performances of various normalization strategies. Principal component analysis (PCA) was used to visualize all samples in the training set ($n = 120$) using genes in GPS ($n = 196$) as variable input. The PCA space was constructed based on (A) reference gene normalized data, or (B) Count Per Total Count (CPTC) normalized data. Samples are colored according to sensitizing properties (red = sensitizers, green = non-sensitizers), and a clear separation between the two groups can be observed along the first PCA component for both normalization strategies. The predictive performance of a model constructed on either normalization strategy was validated using a SVM based LOOCV procedure. A SVM model was trained on all but one sample in the training set, and the left out sample was used to validate the model. The process was repeated until all samples had been used for validation, and the predictive performance of the model was summarized using AUC ROC. The predictive performance of a model based on (C) reference gene normalized data, or (D) CPTC normalized data was identical and estimated to 0.92.

sensitizers and non-sensitizers. However, while cutoff for test set 1 appeared to be well-calibrated, the transcriptional shift previously observed during classification of benchmark set 2, was evident also among samples in test set 2. To verify visual classifications, the SVM model used for prediction of benchmark sets was applied also for prediction of samples in the test sets. However, prior to final classifications, the SVM DVs for each individual sample in the test sets were adjusted against the cutoff threshold for maximum performance for prediction of samples in corresponding benchmark set, to adjust for batch to batch variations between training set and test sets. Calibrated SVM DVs are reported in Supplementary Table 3, and further summarized in combination with GPS expression profiles for each individual sample in Fig. 4. In accordance with visual classifications, a clear discrimination between sensitizers and non-sensitizers are observed in test set 1 (Fig. 4A) and 2 (Fig. 4B), which was verified by ROC AUC values of 0.93 and 0.91, respectively (data not shown). In general, the majority of the sensitizers were assigned higher SVM decision values by the model in comparison to non-sensitizers, which also are reflected by observed differences in expression profiles between most compounds within each group. GARD was able to classify the 29 compounds in

test set 1 and 2 with an accuracy, sensitivity and specificity estimated to 90%, 94% and 83%, indicating a model with good discriminating ability and well calibrated cutoffs for decision making. Supplementary cooper statistics for separate test set are summarized in Table 3.

3.4. Evaluation of alternative supervised machine learning algorithms for classification of samples in independent test sets

To determine if the predictive performance of GPS could be further improved based on the high quality data produced on nCounter, we evaluated a selection of well-described supervised machine learning algorithms for classification, and correlated their performance metrics to SVM classifications. Since model parameters can affect predictive performances, each model was optimized by iteratively evaluating available tuning parameters using cross-validation. Importantly, models were only optimized based on samples in training data in order to prevent overfitting. Optimized models were subsequently trained on samples in training set, calibrated to maximum performance based on classifications of samples in benchmark sets, and applied to classify samples in test sets. In Fig. 5, the calibrated class probabilities for each

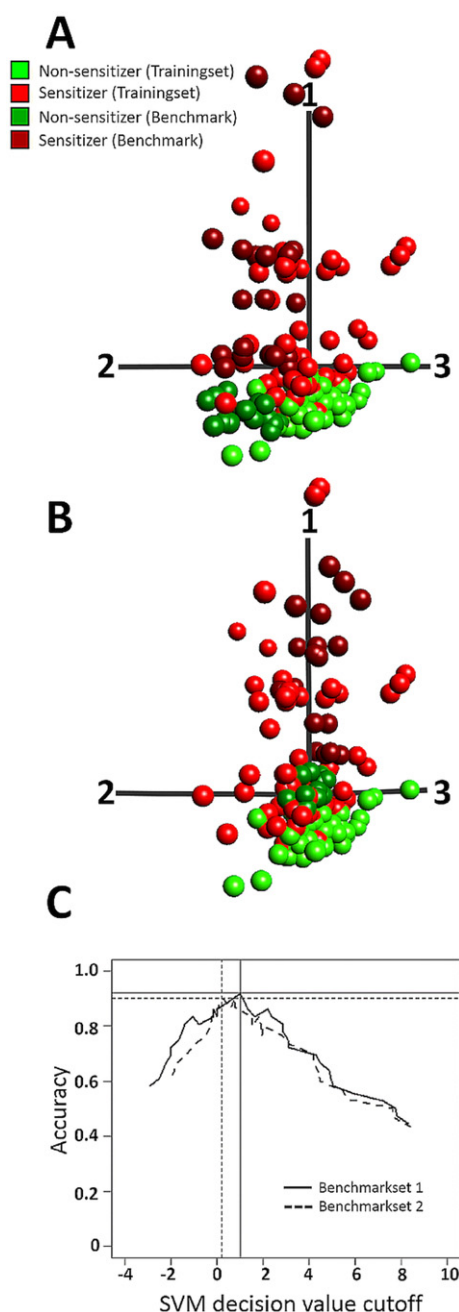


Fig. 2. Calibration of the predictive model using benchmark controls. Unsupervised PCA analysis was used to visually inspect benchmark control samples for batch effects. A three-dimensional PCA space was constructed based on all samples in the training set ($n = 120$), using genes in GPS ($n = 196$) as variable input, and samples were colored according to sensitizing properties (sensitizer = red, non-sensitizer = green). Internal control samples in (A) benchmark set 1 ($n = 36$), and (B) benchmark set 2 ($n = 30$) were plotted into the PCA space constructed by samples in training set, without contributing to PCA components. Control samples were colored according to sensitizing properties in slightly darker colors than samples in the training set. (C) Visual effects were confirmed using a SVM model. The SVM was trained on samples in training set and applied to predict sensitizing properties of samples in each benchmark set. The accuracy of the SVM model was plotted as a function of SVM decision value cutoff (benchmark set 1 = filled line, benchmark set 2 = dashed line).

sample are summarized in series of box- and whisker plots. Based on median values from triplicate stimulations, the accuracies of the PLS model (Fig. 5A for test set 1, and Fig. 5B for test set 2), the KNN model (Fig. 5C and D), and the RF model (Fig. 5E and F) for classification of samples in the independent test sets were estimated to 90%, 93% and

100%, respectively. Additional cooper statistics for each model are summarized in Table 4.

4. Conclusion

Extensive efforts have been invested in the development of mechanistically relevant *in vitro* assays for prediction of chemical skin sensitizers to replace current animal testing. However, to date few assays have been formally validated and none has yet gained full regulatory acceptance to be used as a stand-alone assay (Dumont, Barroso, Matys, Worth, & Casati, 2016). Several attempts have been made to combine information from different assays, each targeting different immunological key mechanisms in sensitization, into an ITS (Bauch et al., 2012; Natsch et al., 2013). Ultimately, the success of such a strategy will be dependent on the ability of the individual ITS components to provide accurate and reliable data.

GARD is an *in vitro* assay for assessment of skin sensitization, addressing the DC activation step in the AOP for skin sensitization. In contrast to assays covering the same key event, such as U-SENSTM (Piroird et al., 2015) and h-CLAT (Ashikaga et al., 2006; Nukada et al., 2011), which rely on single or few markers for classification, the GARD assay is based on measurements of a large set of genes within a genomic biomarker signature, comprising genes involved in immunological relevant pathways associated with skin sensitization. The high informational content enables mechanism-of-action based decisions, well-correlated to current recommendations within toxicity testing (NRC National Research Council, 2007). In this study, we describe modifications to the GARD workflow to better meet future industrial and regulatory demands on reliability, protocol robustness, sample throughput and assay costs. We transferred gene expression measurements from transcriptome-wide microarrays to the more focused NanoString nCounter, measuring only transcripts in the biomarker signature. The platform transfer was validated by generating novel cellular stimulations and evaluating the predictive performance of the GARD assay based on nCounter measurements. A total of 29 compounds were selected as independent test chemicals to estimate model performance. Using internal control samples, so-called benchmarks for calibration of batch effects, compounds in test sets could be classified by the SVM with a sensitivity, specificity and accuracy of 94%, 83% and 90%, reaching similar predictive metrics as the model based on microarray measurements. Of the 16 previously correctly classified compounds, 14 (88%) were correctly predicted by the SVM model based on nCounter measurements in this study, indicating good agreement between platforms. In addition to the predictive performance described for the overall SVM classifications, we also show that the predictive performance of the assay could be improved when applying alternative machine learning models for classification. Changing to a RF model, all of the testing compounds could be accurately assigned to the right class, reaching an accuracy of 100%. Additionally, on the level of individual replicates, the RF model did not have a single misclassification, indicating robustness in predictions. Since limited or insufficient human data was available for most of the chemicals, predictive performances were calculated in relation to LLNA. Exceptions were made for compounds known as false predictions in LLNA, as indicated in Table 2, and for the two chemicals DMTBS and DETBS, where LLNA data was missing. These two chemicals are rubber allergens present in certain disposable rubber gloves, and show positive reactions in human patch testing (Hansson, Ponten, Svedman, & Bergendorff, 2014). Based on a dataset of 139 chemicals, h-CLAT recently reported an accuracy of 78% (Takenouchi et al., 2015). KeratinoSens and DPRA reported similar accuracies (77% and 80% respectively) when challenged with 145 chemicals (Natsch et al., 2013). In this respect, the GARD assay should be considered as highly competitive.

Hazard classification is a first step towards a more complete toxicological safety assessment of chemicals, which would involve an estimate also of relative potency. Similar to binary classifications, a full estimate of potency is anticipated to require a combination of assays targeting

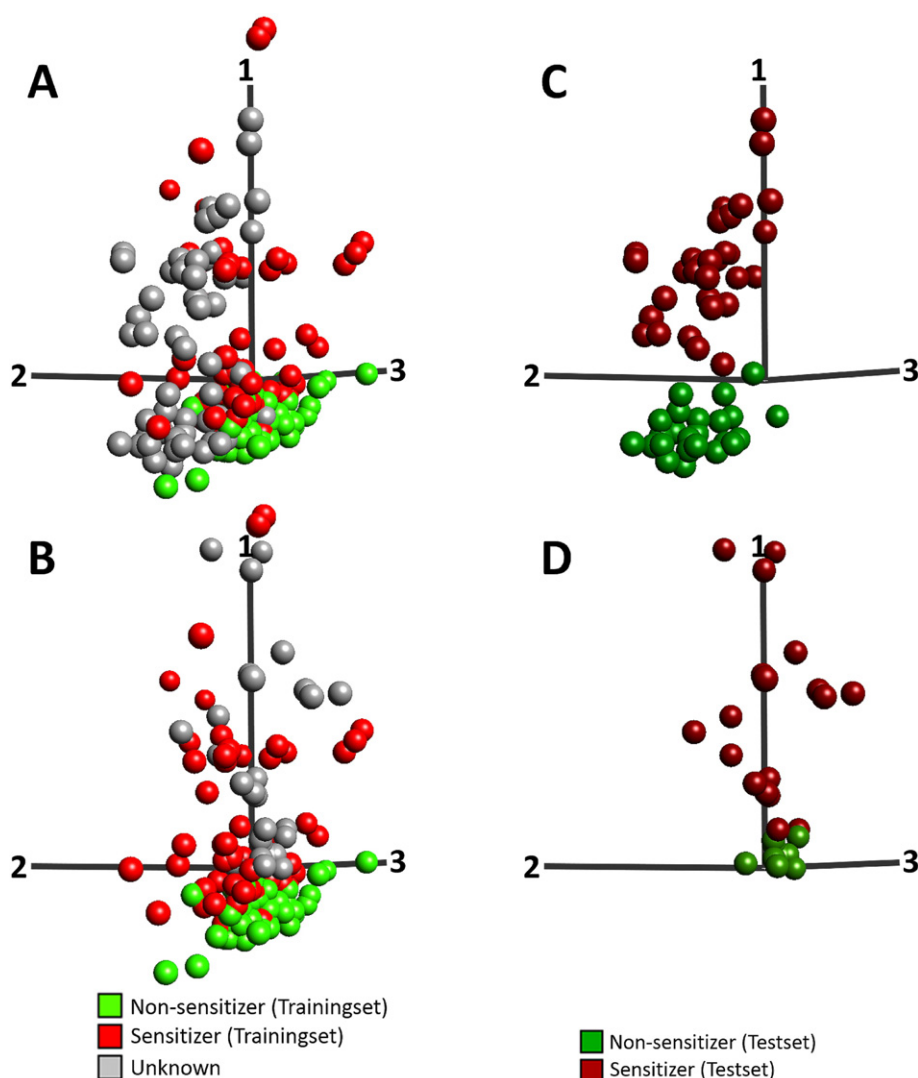


Fig. 3. Visual classifications of samples in independent test sets. A three-dimensional PCA space was constructed by samples in the training set ($n = 120$), based on expression data of the 196 genes in GPS. Samples were colored according to sensitizing properties (sensitizer = red, non-sensitizer = green). Samples in (A) test set 1 ($n = 60$) and (B) test set 2 ($n = 27$) were plotted into the space generated by the training set, without affecting PCA components. Samples in test sets were colored in grey to highlight that they were previously unseen by the model, and not contributed to the representation of the PCA space. True identities of samples in (C) test set 1 and in (D) test set 2 were revealed and colored according to sensitizing properties (sensitizer = dark red, non-sensitizer = dark green), and the training set was removed to facilitate interpretation.

different mechanistic steps in the AOP into an ITS. Of particular interest are such methods based on machine learning algorithms (Tsujita-Inoue, Atobe, Hirota, Ashikaga, & Kouzuki, 2015; Tsujita-Inoue et al., 2014), and Bayesian Networks (Jaworska et al., 2015). The machine learning algorithms used in GARD effectively translate the genomic biomarker signature profiles of each compound into a continuous scale of predictions. It would thus be tempting to evaluate GARD predictions within an ITS for skin sensitizing potency, as the ones suggested above, to elucidate if GARD is able to provide also quantitative assessments.

To improve reliability and reproducibility in predictions, we introduced a novel method to calibrate GARD predictions prior to classification of unknown testing compounds. Since classification of unknown compounds relies on comparison to historical reference data in the training set, test samples will be processed at a different time point in comparison to training samples, thus risking the introduction of technical variations, and complicate predictions. Visual inspections of samples in test sets using a PCA frozen on samples in training set revealed a clear discrimination between sensitizers and non-sensitizers, supporting our

previous claims that the GPS biomarker signature contains genes relevant to the sensitization process. However, discrimination does not intrinsically mean a good calibration, and we observed slight shifts in transcriptional profiles between samples processed in the various batches. To reduce such impacts of batch-to-batch variations, and to avoid having to rerun a complete training set together with new samples, we included a panel of carefully selected samples from the training set, so-called benchmarks, into each test set. These samples with known expected outcomes were used to adjust the cutoff of the classification model to maximize performance prior to evaluating unknown testing compounds. A similar exercise for calibration of SVM outputs have been described by Jiang, Menon, Wang, Kim, & Ohno-Machado (2012). Importantly, the position of the cutoff was only determined based on the benchmark samples, the test compounds were still completely unknown to the model during calibration. However, it is reasonable to assume that such a cutoff will be applicable also to unknown samples processed in same batch. Indeed, the calibrated cutoff improved the accuracy from 79% to 90%, and significantly reduced the number of false

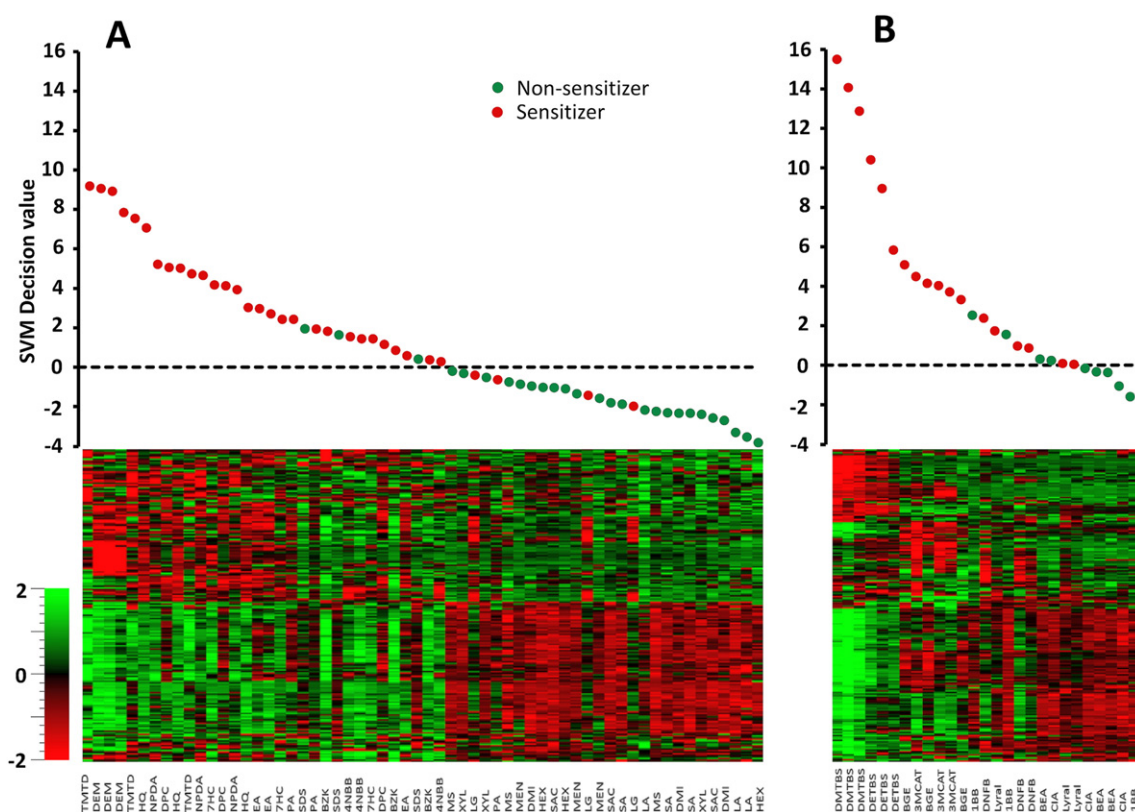


Fig. 4. SVM classifications of samples in independent test sets. A SVM model trained on all samples in the training set ($n = 120$) was applied to predict sensitizing properties of individual replicate samples in (A) test set 1 ($n = 60$) and (B) test set 2 ($n = 27$). Prior to final classification, the SVM DVs for each replicate sample in test sets were first calibrated against the cutoff for maximum accuracy obtained during classification of internal control samples in the corresponding benchmark set (test set 1 = 1.01, test set 2 = 0.23). SVM DVs were sorted in decreasing order from highest to lowest, and colored according to true sensitizing properties (sensitizer = dark red, non-sensitizer = dark green). Dashed lines indicate the cutoff threshold used for classification by the model (SVM DVs > 0 = sensitizer, SVM DVs < 0 = non-sensitizer). On individual replicate level, a total of 87% (76/87) of samples were correctly classified by the model. Heatmaps illustrate expression profiles of transcripts within GPS for each sample in independent test sets.

positive predictions of samples in the test sets. Important to note, changing the cutoff does not change the model or the input variables in the GPS, but it allows for adjusting technical variations in each batch in relation to the biological information in each test set. In addition, since accuracy is a trade-off between sensitivity and specificity, the model can similarly be calibrated to reach a higher sensitivity, which can be of importance to avoid false negative predictions.

Further important actions to facilitate the current workflow of GARD were taken during initial normalization steps. In GARD, novel testing substances may arrive serially in batches, and each sample needs to be compared to historical reference samples. Thus, having to rely on a strategy that requires samples to be processed simultaneously during normalization would be highly impractical, since historical data would need to be renormalized each time new samples arrives. To address the need for a single sample normalization strategy, we evaluated an alternative

method for correction of sample variability, based on Counts per Total Counts (CPTC). The strategy is a modification of Total Count Normalization in RNA-seq (Dillies et al., 2013) and is used to normalize differences in RNA inputs. The suggested normalization strategy performed similar to reference normalization on the samples in the training set, and performed well during predictions of test chemicals.

Finally, the microarray platform will continue to be an important and integrated part of GARD assay development. Measuring complete transcriptomes of cells provides important information in order to study cellular signaling pathways and may be used to extract information relevant to sensitizing potency, as suggested in Albrekt, Johansson, Borje, Borrebaeck, & Lindstedt (2014). The platform has also proven useful for extending the applicability domain of GARD to cover prediction of additional endpoints through identification of separate biomarker signatures, such as chemical respiratory sensitizers (Forreryd, Johansson, Albrekt, Borrebaeck, & Lindstedt, 2015), and recently also for skin sensitizing potency (manuscript under preparation). In conclusion, we here present a streamlined GARD workflow to advance from global biomarker discovery into a reliable and resource-effective format based on the NanoString nCounter platform. When challenged with 29 test compounds, the GARD assay correctly classified 90% based on SVM predictions, validating the predictive relevance of the GPS biomarker signature on an alternative platform. When combining the high quality nCounter data with a powerful RF model for classification, we demonstrated that all 29 test compounds could be accurately predicted. We argue that GARD is well-adapted for routine testing, and that the assay is ready for implementation into an ITS to reflect the DC activation step in the AOP for skin sensitization.

Table 3

Cooper statistics summarizing predictions on individual test sets using the SVM model for classification.

Correctly classified compounds	Test set 1	Test set 2	Total
Sensitizers	10/11	6/6	16/17
Sensitivity (%)	91%	100%	94%
Non-sensitizers	8/9	2/3	10/12
Specificity (%)	89%	67%	83%
Total	18/20	8/9	26/29
Accuracy (%)	90%	89%	90%

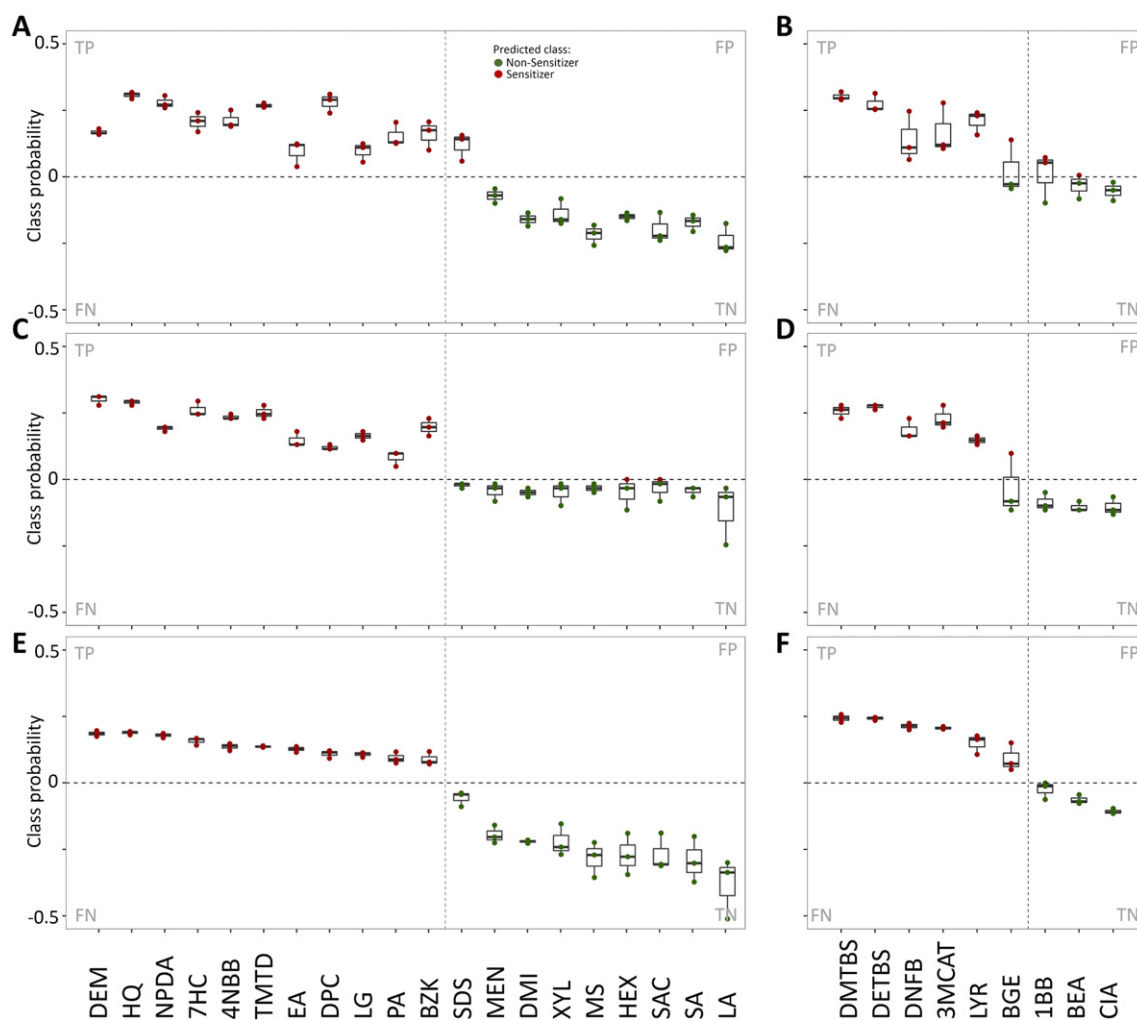


Fig. 5. Classifications of samples in independent test sets using various supervised machine learning models. Machine learning models were trained on all samples in training set ($n = 120$) and applied to predict each individual sample in test sets. Prior to final classifications, the prediction output from each model, the class probability votes, was first calibrated against the cutoff for maximum accuracy obtained during classification of internal control samples in the corresponding benchmark set. Calibrated class probability votes for triplicate samples of each chemical were summarized in series of box and whisker plots for the (A) partial least squares (PLS) model on test set 1 and (B) test set 2, the (C) K-nearest neighbor (KNN) model on test set 1 and (D) test set 2, and the (E) random forest (RF) model on test set 1 and (F) test set 2. Samples were colored based on classifications provided by each model (sensitizer = dark red, non-sensitizer = dark green). The horizontal dashed line in each subplot represents the cutoff threshold used by models for classification as sensitizers (class probability votes > 0), or non-sensitizers (class probability votes < 0), and the vertical dashed line marks the separation between true sensitizers and true non-sensitizers (left = sensitizers, right = non-sensitizers). Using median value of replicates for classification of a certain chemical, the accuracy of the PLS, KNN and the RF models was estimated to 90%, 93% and 100%, respectively. TP = true positive, FP = false positive, TN = true negative, FN = false negative.

Conflicts of interest

All authors state no conflict of interest.

Transparency document

The [Transparency document](#) associated with this article can be found, in the online version.

Table 4

Cooper statistics summarizing predictions on test sets using PLS, KNN and RF for classifications.

Correctly classified compounds	PLS	KNN	RF
Sensitizers	16/17	16/17	17/17
Sensitivity (%)	94%	94%	100%
Non-sensitizers	10/12	11/12	12/12
Specificity (%)	83%	92%	100%
Total	26/29	27/29	29/29
Accuracy (%)	90%	93%	100%

Acknowledgements

This work was supported by grants from the Swedish Research Council VR (521-2014-2286), Swedish Foundation for Strategic Research (IB13-0118), and the Faculty of Engineering (LTH). We would like to thank Associate Professors Ola Bergendorff and Malin Engfeldt for providing the chemical compounds DETBS and DMTBS.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at <http://dx.doi.org/10.1016/j.tiv.2016.09.013>.

References

- Akhavan, A., Cohen, S.R., 2003. The relationship between atopic dermatitis and contact dermatitis. *Clin. Dermatol.* 21 (2), 158–162.
- Albrekt, A.S., Johansson, H., Borje, A., Borrebaeck, C., Lindstedt, M., 2014. Skin sensitizers differentially regulate signaling pathways in MUTZ-3 cells in relation to their

- individual potency. *BMC Pharmacol. Toxicol.* 15 (5). <http://dx.doi.org/10.1186/2050-6511-15-5>.
- Andersen, C.L., Jensen, J.L., Orntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64 (15), 5245–5250.
- Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H., 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 (5), 767–773.
- Basketter, D.A., Evans, P., Fielder, R.J., Gerberick, G.F., Dearman, R.J., Kimber, I., 2002. Local lymph node assay - validation, conduct and use in practice. *Food Chem. Toxicol.* 40 (5), 593–598.
- Bauch, C., Kolle, S.N., Ramirez, T., Eltze, T., Fabian, E., Mehling, A., Teubner, W., van Ravenzwaay, B., Landsiedel, R., 2012. Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. *Regul. Toxicol. Pharmacol.* 63 (3), 489–504.
- Bergendorff, O., Persson, C., Ludtke, A., Hansson, C., 2007. Chemical changes in rubber allergens during vulcanization. *Contact Dermatitis* 57 (3), 152–157.
- Breiman, L., 2001. Random Forest. *Mach. Learn.* 45, 5–32.
- Cooper II, J.A., Saracci, R., Cole, P., 1979. Describing the validity of carcinogen screening tests. *Br. J. Cancer* 39 (1), 87–89.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297.
- DeGroot, A., 2008. Test Concentrations and Vehicles for 4350 Chemicals. 3rd ed. Acdegroot Publishing, Wapserveen, The Netherlands.
- Dickel, H., Kuss, O., Blesius, C.R., Schmidt, A., Diepgen, T.L., 2001. Occupational skin diseases in Northern Bavaria between 1990 and 1999: a population-based study. *Br. J. Dermatol.* 145 (3), 453–462.
- Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., et al., 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.* 14 (6), 671–683.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., 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 May 19, 2016).
- Draghici, S., Khatri, P., Eklund, A.C., Szallasi, Z., 2006. Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet.* 22 (2), 101–109.
- Dumont, C., Barroso, J., Matys, I., Worth, A., Casati, S., 2016. Analysis of the local lymph node assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. *Toxicol. In Vitro* 34, 220–228.
- EU, 2006. Regulation (EC) no 1907/2006 of the European Parliament and of the council 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.
- 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 L* 342, 1–59.
- Fonacier, L.S., Dreskin, S.C., Leung, D.Y., 2010. Allergic skin diseases. *J. Allergy Clin. Immunol.* 125 (2 Suppl 2), S138–S149.
- Forreryd, A., Johansson, H., Albrekt, A.S., Borrebaeck, C.A., Lindstedt, M., 2015. Prediction of chemical respiratory sensitizers using GARD, a novel in vitro assay based on a genomic biomarker signature. *PLoS One* 10 (3), e0118808. <http://dx.doi.org/10.1371/journal.pone.0118808>.
- 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. <http://dx.doi.org/10.1186/1471-2164-15-379>.
- Geier, J., Lessmann, H., Uter, W., Schnuch, A., 2003. Occupational rubber glove allergy: results of the information network of departments of dermatology (IVDK), 1995–2001. *Contact Dermatitis* 48 (1), 39–44.
- Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., James, J.J., Maysuria, M., Mitton, J.D., Oliveri, P., Osborn, J.L., Peng, T., Ratcliffe, A.L., Webster, P.J., Davidson, E.H., Hood, L., Dimitrov, K., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* 26 (3), 317–325.
- Gerberick, G.F., Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A., 2005. Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis* 16 (4), 157–202.
- Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., Lepoittevin, J.P., 2004. Development of a peptide reactivity assay for screening contact allergens. *Toxicol. Sci.* 81 (2), 332–343.
- Haneke, K.E., Tice, R.R., Carson, B.L., Margolin, B.H., 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 (3), 274–286.
- Hansson, C., Ponten, A., Svedman, C., Bergendorff, O., 2014. Reaction profile in patch testing with allergens formed during vulcanization of rubber. *Contact Dermatitis* 70 (5), 300–308.
- Jaworska, J.S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., Miyazawa, M., 2015. Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. *Arch. Toxicol.* 89 (12), 2355–2383.
- Jiang, X., Menon, A., Wang, S., Kim, J., Ohno-Machado, L., 2012. Doubly optimized calibrated support vector machine (DOC-SVM): an algorithm for joint optimization of discrimination and calibration. *PLoS One* 7 (11), e48823. <http://dx.doi.org/10.1371/journal.pone.0048823>.
- Johansson, H., Albrekt, A.S., Borrebaeck, C.A., Lindstedt, M., 2013. The GARD assay for assessment of chemical skin sensitizers. *Toxicol. In Vitro* 27 (3), 1163–1169. <http://dx.doi.org/10.1016/j.tiv.2012.05.019>.
- Johansson, H., Lindstedt, M., Albrekt, A.S., 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. <http://dx.doi.org/10.1186/1471-2164-12-399>.
- Johansson, H., Rydnert, F., Kuhn, J., Schepky, A., Borrebaeck, C., Lindstedt, M., 2014. Genomic allergen rapid detection in-house validation—a proof of concept. *Toxicol. Sci.* 139 (2), 362–370.
- Kimber, I., Basketter, D.A., Gerberick, G.F., Dearman, R.J., 2002. Allergic contact dermatitis. *Int. Immunopharmacol.* 2 (2–3), 201–211.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, R., Mayer, Z., Kenkel, B., Benesty, M., et al., 2016. Caret: Classification and Regression Training. R Package Version 6. <http://CRAN.R-project.org/package=caret> (Accessed May 19, 2016).
- Lasko, T.A., Bhagwat, J.G., Zou, K.H., Ohno-Machado, L., 2005. The use of receiver operating characteristic curves in biomedical informatics. *J. Biomed. Inform.* 38 (5), 404–415.
- Liang, P., 2007. MAQC papers over the cracks. *Nat. Biotechnol.* 25 (1), 27–28 (author reply 28–9).
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2 (3), 18–22.
- Lunder, T., Kansky, A., 2000. Increase in contact allergy to fragrances: patch-test results 1989–1998. *Contact Dermatitis* 43 (2), 107–109.
- Mevik, B.W., Hovde Liland, R.K., 2015. pls: Partial Least Squares and Principal Component Regression. R Package Version 2.5. <http://CRAN.R-project.org/package=pls> (Accessed May 19, 2016).
- Murphy, D., 2002. Gene expression studies using microarrays: principles, problems, and prospects. *Adv. Physiol. Educ.* 26 (1–4), 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 (2), 284–292.
- Natsch, A., Ryan, C.A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., Kern, P., 2013. A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalence. *J. Appl. Toxicol.* 33 (11), 1337–1352.
- Nguyen, S.H., Dang, T.P., MacPherson, C., Maibach, H., Maibach, H.I., 2008. Prevalence of patch test results from 1970 to 2002 in a multi-Centre population in North America (NACDG). *Contact Dermatitis* 58 (2), 101–106.
- NRC National Research Council, 2007. Toxicity Testing in the 21st Century: A Vision and a Strategy. National Academies Press, Washington, DC.
- Nukada, Y., Ashikaga, T., Sakaguchi, H., Sono, S., Mugita, N., Hirota, M., Miyazawa, M., Ito, Y., Sasa, H., Nishiyama, N., 2011. Predictive performance for human skin sensitizing potential of the human cell line activation test (h-CLAT). *Contact Dermatitis* 65 (6), 343–353.
- 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.
- 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., 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 (5), 763–781.
- Piroid, C., Ovigne, J.M., Rousset, F., Martinozzi-Teissier, S., Gomes, C., Cotovio, J., Alepee, N., 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 (5), 901–916.
- R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org> (Accessed May 19, 2016).
- Reisinger, K., Hoffmann, S., Alepee, N., Ashikaga, T., Barroso, J., Elcombe, C., Gellatly, N., Galbiati, V., Gibbs, S., Groux, H., et al., 2015. Systematic evaluation of non-animal test methods for skin sensitisation safety assessment. *Toxicol. In Vitro* 29 (1), 259–270.
- Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCr: visualizing classifier performance in R. *Bioinformatics* 21 (20), 3940–3941.
- Takenouchi, O., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., et al., 2015. Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *J. Appl. Toxicol.* 35 (11), 1318–1332.
- Thyssen, J.P., Linneberg, A., Menne, T., Johansen, J.D., 2007. The epidemiology of contact allergy in the general population—prevalence and main findings. *Contact Dermatitis* 57 (5), 287–299.
- Tsujita-Inoue, K., Atobe, T., Hirota, M., Ashikaga, T., Kouzuki, H., 2015. In silico risk assessment for skin sensitization using artificial neural network analysis. *J. Toxicol. Sci.* 40 (2), 193–209.
- Tsujita-Inoue, K., Hirota, M., Ashikaga, T., Atobe, T., Kouzuki, H., Aiba, S., 2014. Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays. *Toxicol. In Vitro* 28 (4), 626–639.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paep, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3 (7) Research0034.
- Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer, New York <http://ggplot2.org> (Accessed May 19, 2016).

Table S1. GPS prediction signature

Gene Symbol	Gene Title	Accession	Affymetrix Transcript cluster ID	NanoString Probe ID (NSID)
ABAT	4-aminobutyrate aminotransferase	18	7993126	NM_020686.5:268
ABHD5	abhydrolase domain containing 5	51099	8079153	NM_016006.4:910
ACER2	alkaline ceramidase 2	340485	8154563	NM_001010887.2:1306
ACLY	ATP citrate lyase	47	8015460	NM_001096.2:3990
ACTR10	actin-related protein 10 homolog (S. cerevisiae)	55860	7974587	NM_018477.2:1016
ADAM20	ADAM metallopeptidase domain 20	8748	7979927	NM_003814.4:1420
ALDH18A1	aldehyde dehydrogenase 18 fam., member A1	5832	7935230	NM_001017423.1:2617
ALDH1B1	aldehyde dehydrogenase 1 fam., member B1	219	8155327	NM_000692.3:1255
ANAPC1	anaphase promoting complex subunit 1	64682	8043349	NM_022662.3:7202
ANAPC5	anaphase promoting complex subunit 5	51433	7967149	NM_016237.4:1444
ANKRA2	ankyrin repeat, fam. A (RFXANK-like), 2	57763	8112596	NM_023039.4:741
ARFGAP3	ADP-ribosylation factor GTPase activating protein 3	26286	8076515	NM_001142293.1:2362
ARHGAP9	Rho GTPase activating protein 9	64333	7964436	BC006107.1:1808
ASB7	ankyrin repeat and SOCS box-containing 7	140460	7986433	NM_024708.3:1280
ATP6V0D1//A TP6V0D1	ATPase, H+ transporting, lysosomal 38 kDa, V0 subunit d1//ATPase, H+ transporting, lysosomal 38 kDa, V0 subunit d1	9114//911 4	8002041	NM_004691.4:1101
ATP6V0E1	ATPase, H+ transporting, lysosomal 9 kDa, V0 subunit e1	8992	8110022	NM_003945.3:617
ATP6V1H	ATPase, H+ transporting, lysosomal 50/57 kDa, V1 subunit H	51606	8150797	NM_213620.2:1095
BCL7A	B-cell CLL/lymphoma 7A	605	7959354	NM_001024808.1:594
BIN2	bridging integrator 2	51411	7963289	NM_016293.2:855
BLMH	bleomycin hydrolase	642	8014008	NM_000386.3:2088
BXDC1//RPF2	brix domain containing 1//ribosome production factor 2 homolog (S. cerevisiae)	84154//84 154	8062211	ENST00000424137.1:2 88
C11orf61	chromosome 11 open reading frame 61	79684	7952445	NM_024631.2:1622
C11orf67//INT S4	chromosome 11 open reading frame 67//integrator complex subunit 4	28971//92 105	7942783	NM_024684.2:110
C12orf57	chromosome 12 open reading frame 57	113246	7953564	NM_138425.2:152
C13orf18	chromosome 13 open reading frame 18	80183	7971486	NM_025113.3:648
C15orf24	chromosome 15 open reading frame 24	56851	7987172	NM_020154.2:147
C19orf46//AL KBH6	chromosome 19 open reading frame 46//alkB, alkylation repair homolog 6 (E. coli)	163183//8 4964	8036242	NM_032878.3:838
C19orf54	chromosome 19 open reading frame 54	284325	8036956	NM_198476.3:1349
C1orf174	chromosome 1 open reading frame 174	339448	7911897	NM_207356.2:1372
C1orf183	chromosome 1 open reading frame 183	55924	7918552	NM_019099.4:410
C20orf111	chromosome 20 open reading frame 111	51526	8066402	NM_016470.7:1035
C20orf24	chromosome 20 open reading frame 24	55969	8062326	NR_026562.1:645
C3orf62//USP4	chromosome 3 open reading frame 62//ubiquitin specific peptidase 4 (proto-oncogene)	375341//7 375	8087374	ENST00000343010.1:9 42
C9orf89	chromosome 9 open reading frame 89	84270	8156404	NM_032310.3:204
CARM1	coactivator-associated arginine methyltransferase 1	10498	8025766	NM_199141.1:2352
CD33	CD33 molecule	945	8030804	NM_001082618.1:786
CD86	CD86 molecule	942	8082035	NM_175862.4:289
CD93	CD93 molecule	22918	8065359	NM_012072.3:4270
COX7A2L	cytochrome c oxidase subunit VIIa polypeptide 2 like	9167	8051777	NM_004718.2:635
CRHBP	corticotropin releasing hormone binding protein	1393	8106418	NM_001882.3:1140
CSGALNACT 2	chondroitin sulfate N- acetylgalactosaminyltransferase 2	55454	7927146	NM_018590.4:647
CYP51A1	cytochrome P450, fam. 51, subfam. A, polypeptide 1	1595	8140864	NM_000786.3:240
DDRKG1	DDRKG domain containing 1	65992	8064601	NM_023935.1:800
DDX19A	DEAD (Asp-Glu-Ala-As) box polypeptide 19A	55308	7997059	NM_018332.3:2236
DDX21	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	9188	7927936	NM_004728.3:338
DHCR24	24-dehydrocholesterol reductase	1718	7916432	NM_014762.3:1385
DHCR7	7-dehydrocholesterol reductase	1717	7950067	NM_001360.2:2352
DHX33	DEAH (Asp-Glu-Ala-His) box polypeptide 33	56919	8011861	NM_001199699.1:2550
DNAJB4	DnaJ (Hsp40) homolog, subfam. B, member 4	11080	7902512	NM_007034.3:949
DNAJB9	DnaJ (Hsp40) homolog, subfam. B, member 9	4189	8135480	NM_012328.2:274

DNAJC5	DnaJ (Hsp40) homolog, subfam. C, member 5	80331	8064208	NM_025219.2:1510
DNAJC9	DnaJ (Hsp40) homolog, subfam. C, member 9	23234	7934320	NM_015190.3:42
DRAM2//CEP T1	DNA-damage regulated autophagy modulator 2//choline/ethanolamine phosphotransferase 1	128338//1 0390	7918474	NM_178454.4:628
DTD1	D-tyrosyl-tRNA deacylase 1 homolog (S. cerevisiae)	92675	8061211	NM_080820.4:1085
EDEM2	ER degradation enhancer, mannosidase alpha-like 2	55741	8065855	NM_018217.2:1366
EVI2B	ecotropic viral integration site 2B	2124	8014063	NM_006495.3:467
FAM36A//NC RNA00201	fam. with sequence similarity 36, member A	116228//2 84702	7911085	NM_198076.4:155
FAM86A	fam. with sequence similarity 86, member A	196483	7999304	NM_201400.3:654
FAS	Fas (TNF receptor superfam., member 6)	355	7929032	NM_000043.4:398
FASN	fatty acid synthase	2194	8019392	NM_004104.4:1450
FBXO10//TO MM5	F-box protein 10//translocase of outer mitochondrial membrane 5 homolog (yeast)	26267//40 1505	8161229	NM_012166.2:651
FDPSP2	farnesyl diphosphate synthase pseudogene 2	619190	8140443	NR_003262.1:388
FDXR	ferredoxin reductase	2232	8018236	NM_004110.4:1758
FOXO4	forkhead box O4	4303	8168205	NM_005938.3:1252
FTH1P5	ferritin, heavy polypeptide 1 pseudogene 5	2509	8126948	J04755.1:308
FUCA2	fucosidase, alpha-L- 2, plasma	2519	8129974	NM_032020.4:868
GAS2L3	growth arrest-specific 2 like 3	283431	7957850	NM_174942.1:1315
GDAP2	ganglioside induced differentiation associated protein 2	54834	7918955	NM_001135589.2:957
GDF11	growth differentiation factor 11	10220	7956026	NM_005811.3:738
GLRX	glutaredoxin (thioltransferase)	2745	8113214	NM_002064.2:309
GNL3L	guanine nucleotide binding protein-like 3 (nucleolar)-like	54552	8167797	NM_001184819.1:3915
GNPNAT1	glucosamine-phosphate N-acetyltransferase 1	64841	7979196	NM_198066.3:2056
GSR	glutathione reductase	2936	8150112	NM_000637.3:1628
GTF2IRD2//G TF2IRD2B	GTF2I repeat domain containing 2//GTF2I repeat domain containing 2B	84163//38 9524	8133549	GPS_8133549b.1
GTF2IRD2//G TF2IRD2B	GTF2I repeat domain containing 2//GTF2I repeat domain containing 2B	84163//38 9524	8140170	GPS_8140170.1
GTF3C2	general transcription factor IIIC, polypeptide 2, beta 110 kDa	2976	8051075	NM_001521.2:3555
HBP1//COG5	HMG-box transcription factor 1//component of oligomeric golgi complex 5	26959//10 466	8135392	NM_012257.3:1330
HIST1H1C	histone cluster 1, H1c	3006	8124397	NM_005319.3:401
HIST1H1E	histone cluster 1, H1e	3008	8117377	NM_005321.2:172
HIST1H2AE	histone cluster 1, H2ae	3012	8117408	NM_021052.2:455
HIST1H2BE	histone cluster 1, H2be	8344	8117389	NM_003523.2:321
HIST1H3G	histone cluster 1, H3g	8355	8124440	NM_003534.2:348
HIST1H3J	histone cluster 1, H3j	8356	8124537	NM_003535.2:364
HIST1H4A	histone cluster 1, H4a	8359	8117334	NM_003538.3:47
HIST- 2H2AB//2H2A C//2H2AA3//2 H2AA4	histone cluster 2, H2ab//histone cluster 2, H2ac//histone cluster 2, H2aa3//histone cluster 2, H2aa4	8338//833 7//723790	7905079	BC001629.1:306
HIST- 2H2AB//2H2A C//2H2AA3//2 H2AA4	histone cluster 2, H2ab//histone cluster 2, H2ac//histone cluster 2, H2aa3//histone cluster 2, H2aa4	8338//833 7//723790	7919619	BC001629.1:2
HIST- 2H2BF//2H2B E//2H2BA	histone cluster 2, H2bf//histone cluster 2, H2be//histone cluster 2, H2ba	440689//8 349//3378 75	7919606	NM_001024599.2:1305
HMGB3	high-mobility group box 3	3149	8170468	NM_005342.3:2332
HMGCR//HM GCR	3-hydroxy-3-methylglutaryl-Coenzyme A reductase//3-hydroxy-3-methylglutaryl-CoA reductase	3156//315 6	8106280	NM_000859.2:3634
HMGCS1//HM GCS1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)//3-hydroxy-3-methylglutaryl-CoA synthase 1 (soluble)	3157//315 7	8111941	NM_002130.7:335
HMOX1	heme oxygenase (decycling) 1	3162	8072678	NM_002133.2:866
HNRNPL	heterogeneous nuclear ribonucleoprotein L	3191	8036613	NM_001533.2:1382
IRS2	insulin receptor substrate 2	8660	7972745	NM_003749.2:775
ISCU	iron-sulfur cluster scaffold homolog (E. coli)	23479	7958414	NM_014301.3:733
ISG20L2	interferon stimulated exonuclease gene 20 kDa-like	81875	7921110	NM_030980.1:1492

	2			
KCNE3	potassium voltage-gated channel, Isk-related fam., member 3	10008	7950409	NM_005472.4:2528
KGFLP1//FGF7//KGFLP2//F LJ20444	keratinocyte growth factor-like protein 1//fibroblast growth factor 7 (keratinocyte growth factor)//keratinocyte growth factor-like protein 2//hypothetical protein FLJ20444	387628//252//654466//403323	8155530	NR_003674.2:542
LPAR1	lysophosphatidic acid receptor 1	1902	8163257	NM_001401.3:1184
LRPPRC	leucine-rich PPR-motif containing	10128	8051882	NM_133259.3:4710
LY96	lymphocyte antigen 96	23643	8146934	NM_015364.2:360
MAP2K1//SNAPC5	mitogen-activated protein kinase kinase 1//small nuclear RNA activating complex, polypeptide 5, 19 kDa	5604//10302	7984319	NM_002755.3:640
MAPK13	mitogen-activated protein kinase 13	5603	8119016	NM_002754.4:1004
METTL2A	methyltransferase like 2A	339175	8009008	NM_181725.3:298
MGST3	microsomal glutathione S-transferase 3	4259	7906978	NM_004528.2:195
MRPL30	mitochondrial ribosomal protein L30	51263	8043848	NM_145212.3:2278
MRPL4	mitochondrial ribosomal protein L4	51073	8025586	NM_015956.2:766
MRPS17//GBAS//ZNF713	mitochondrial ribosomal protein S17//glioblastoma amplified sequence//zinc finger protein 713	51373//2631//349075	8132922	NM_015969.2:238
MTPAP//LOC729668	mitochondrial poly(A) polymerase	55149//729668	7932834	NM_017442.3:656
MTR	5-methyltetrahydrofolate-homocysteine methyltransferase	4548	7910752	NM_000254.2:6816
NBR1	neighbor of BRCA1 gene 1	4077	8007471	NM_031858.2:602
NIP7	nuclear import 7 homolog (S. cerevisiae)	51388	7996934	NM_016101.4:880
NLRP12	NLR fam., pyrin domain containing 12	91662	8039096	NM_033297.2:2097
NOL6	nucleolar protein fam. 6 (RNA-associated)	65083	8160682	NM_022917.4:4362
NQO1	NAD(P)H dehydrogenase, quinone 1	1728	8002303	NM_000903.2:790
NRBP1	nuclear receptor binding protein 1	29959	8040927	NM_013392.2:1681
NUBPL	nucleotide binding protein-like	80224	7973826	NM_025152.2:374
NUDT14	nudix (nucleoside diphosphate linked moiety X)-type motif 14	256281	7981566	NM_177533.4:712
NUFIP1	nuclear fragile × mental retardation protein interacting protein 1	26747	7971361	NM_012345.2:1005
NUP153	nucleoporin 153 kDa	9972	8124059	NM_001278209.1:4915
OR5B21	olfactory receptor, fam. 5, subfam. B, member 21	219968	7948330	NM_001005218.1:292
PASK	PAS domain containing serine/threonine kinase	23178	8060205	NM_015148.3:2619
PAWR	PRKC, apoptosis, WT1, regulator	5074	7965112	NM_002583.2:824
PDAP1	PDGFA associated protein 1	11333	8141273	NM_014891.6:364
PDE1B	phosphodiesterase 1B, calmodulin-dependent	5153	7955943	NM_000924.3:498
PFAS	phosphoribosylformylglycinamide synthase	5198	8004804	NM_012393.2:4655
PHLDA3	pleckstrin homology-like domain, fam. A, member 3	23612	7923372	NM_012396.3:532
PIK3AP1	phosphoinositide-3-kinase adaptor protein 1	118788	7935337	NM_152309.2:598
PINK1	PTEN induced putative kinase 1	65018	7898663	NM_032409.2:1610
PMM2	phosphomannomutase 2	5373	7993148	NM_000303.2:36
PNO1	partner of NOB1 homolog (S. cerevisiae)	56902	8042381	NM_020143.2:875
POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25 kDa	5434	8032149	NM_002695.3:340
POLR3E	polymerase (RNA) III (DNA directed) polypeptide E (80 kD)	55718	7993973	NM_018119.3:3054
PPM1D//PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform//protein phosphatase, Mg2+/Mn2+ dependent, 1D	8493//8493	8008922	NM_003620.3:2783
PREX1	phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 1	57580	8066848	NM_020820.3:6075
PSTPIP1	proline-serine-threonine phosphatase interacting protein 1	9051	7985099	NM_003978.3:1752
PTMA	prothymosin, alpha	5757	7954006	OTTHUMT00000400136.1:189
PTMA	prothymosin, alpha	5757	7961022	NR_036466.1:476
RAB33B	RAB33B, member RAS oncogene fam.	83452	8097507	NM_031296.1:825
RENBP	renin binding protein	5973	8175933	NM_002910.5:480
RFC2	replication factor C (activator 1) 2, 40 kDa	5982	8140151	NM_181471.2:921

RNASEH1	ribonuclease H1	246243	8050079	NM_002936.4:1003
RNF146	ring finger protein 146	81847	8121927	NM_030963.2:510
RNF24	ring finger protein 24	11237	8064766	NM_007219.2:955
RNF26	ring finger protein 26	79102	7944510	NM_032015.3:2275
RPSA//SNOR A62	ribosomal protein SA//small nucleolar RNA, H/ACA box 62	3921//604 4	8078918	NM_002295.4:160
RPUSD2	RNA pseudouridylate synthase domain containing 2	27079	7982753	NM_152260.2:462
RRP12	ribosomal RNA processing 12 homolog (S. cerevisiae)	23223	7935425	NM_001145114.1:112
RXRA	retinoid × receptor, alpha	6256	8159127	NM_002957.5:3940
SCARB2	scavenger receptor class B, member 2	950	8101158	NM_005506.3:3015
SERBP1	SERPINE1 mRNA binding protein 1	26135	7916836	NM_001018068.1:4610
SFPQ//SFPQ	splicing factor proline/glutamine-rich (polypyrimidine tract binding protein associated)//splicing factor proline/glutamine-rich	6421//642 1	7914791	ENST00000357214.1:2798
SLC25A32//D CAF13	solute carrier fam. 25, member 32//DDB1 and CUL4 associated factor 13	81034//25 879	8152255	NM_030780.4:1323
SLC35B3	solute carrier fam. 35, member B3	51000	8123825	NM_001142540.1:501
SLC37A4	solute carrier fam. 37 (glucose-6-phosphate transporter), member 4	2542	7952132	NM_001164277.1:1248
SLC5A6	solute carrier fam. 5 (sodium-dependent vitamin transporter), member 6	8884	8051030	NM_021095.2:548
SMPD4	sphingomyelin phosphodiesterase 4, neutral membrane (neutral sphingomyelinase-3)	55627	8055183	NM_017751.4:2258
SNHG1//SNO RD26	small nucleolar RNA host gene 1 (non-protein coding)//small nucleolar RNA, C/D box 26	23642//93 02	7948908	AK095849.1:268
SNHG12	small nucleolar RNA host gene 12 (non-protein coding)	85028	7914202	NR_024127.1:560
SNORA45	small nucleolar RNA, H/ACA box 45	677826	7938293	NR_002977.1:0
SNX27	sorting nexin fam. member 27	81609	7905444	NM_030918.5:1770
SPNS2//MYB BP1A	spinster homolog 2 (Drosophila)//MYB binding protein (P160) 1a	124976//1 0514	8011640	NM_014520.3:2644
SPRY2	sprouty homolog 2 (Drosophila)	10253	7972217	NM_005842.2:1481
SQLE	squalene epoxidase	6713	8148280	NM_003129.3:250
SREBF2	sterol regulatory element binding transcription factor 2	6721	8073522	NM_004599.3:3905
ST3GAL6	ST3 beta-galactoside alpha-2,3-sialyltransferase 6	10402	8081219	NM_001271148.1:1000
STK17B	serine/threonine kinase 17b	9262	8057887	NM_004226.3:821
TAPT1	transmembrane anterior posterior transformation 1	202018	8099506	NM_153365.2:1777
TAS2R5	taste receptor, type 2, member 5	54429	8136647	NM_018980.2:248
TBCEL	tubulin folding cofactor E-like	219899	7944623	NM_152715.3:1350
TCTN2	tectonic fam. member 2	79867	7959638	NM_001143850.2:75
TLR6	toll-like receptor 6	10333	8099841	NM_006068.4:261
TMEM150B	transmembrane protein 150B	284417	8039453	NM_001085488.2:723
TMEM55A	transmembrane protein 55A	55529	8151756	NM_018710.2:957
TMEM59	transmembrane protein 59	9528	7916372	NM_004872.3:316
TMEM97	transmembrane protein 97	27346	8005839	NM_014573.2:446
TNFRSF10C	tumor necrosis factor receptor superfam., member 10c, decoy without an intracellular domain	8794	8145244	GPS_8145244b.1:46
TOMM34	translocase of outer mitochondrial membrane 34	10953	8066461	NM_006809.4:1436
TOMM40	translocase of outer mitochondrial membrane 40 homolog (yeast)	10452	8029521	NM_001128916.1:1584
TP53I3	tumor protein p53 inducible protein 3	9540	8050702	NM_001206802.2:16
TP53INP1	tumor protein p53 inducible nuclear protein 1	94241	8151890	NM_001135733.1:700
TWF2//TLR9	twinfilin, actin-binding protein, homolog 2 (Drosophila)//toll-like receptor 9	11344//54 106	8087860	NM_007284.3:284
TXNRD1	thioredoxin reductase 1	7296	7958174	NM_001093771.2:1018
UFC1	ubiquitin-fold modifier conjugating enzyme 1	51506	7906662	NM_016406.3:911
USP10	ubiquitin specific peptidase 10	9100	7997633	NM_001272075.1:1660
VAMP3	vesicle-associated membrane protein 3 (cellubrevin)	9341	7897370	NM_004781.3:1875
VARS	valyl-tRNA synthetase	7407	8125091	NM_006295.2:680
VARS	valyl-tRNA synthetase	7407	8178609	BC12808.2:3946
VPS37A	vacuolar protein sorting 37 homolog A (S. cerevisiae)	137492	8144774	NM_152415.2:920
ZNF211	zinc finger protein 211	10520	8031792	NM_001265600.1:714

ZNF223	zinc finger protein 223	7766	8029360	NM_013361.4:690
ZNF561	zinc finger protein 561	93134	8033795	NM_152289.2:832
ZNF79	zinc finger protein 79	7633	8158022	NM_007135.2:101
---	---	---	7910385	ENST00000466086.1:182
---	---	---	7966223	ENST00000363404.1:5
---	---	---	7979694	ENST00000445822.1:110
---	---	---	8130495	ENST00000538672.1:63
---	---	---	8180237	a
---	---	---	8180268	a
---	---	---	8180417	a
---	---	---	7946567	a

^a Indicates the four targets from the GPS signature which could not be matched to appropriate Nanostring probe IDs and included in the custom CodeSet for nCounter measurements.

Table S2. Candidate reference genes for normalization of data included in the custom CodeSet.

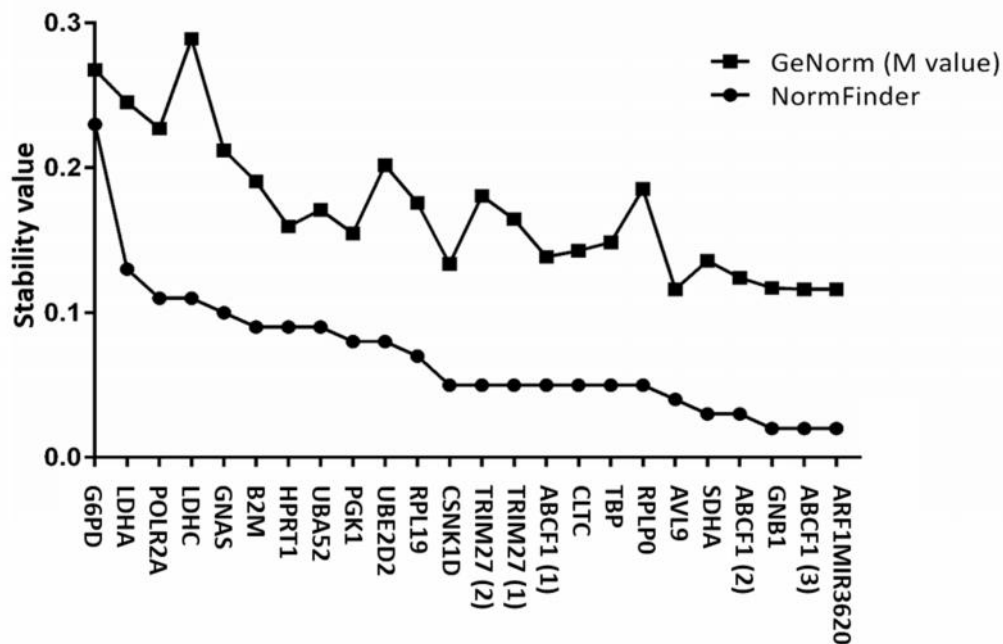
Gene Symbol	Gene Title	Accession	Affymetrix cluster ID	NanoString
ABCF1 (2)	ATP-binding cassette, sub-family F (GCN20), member 1	NM_001090.2	8179112	NM_001090.2:796
ABCF1 (1)	ATP-binding cassette, sub-family F (GCN20), member 1	BC034488.2	8177797	BC034488.2:836
ABCF1 (3)	ATP-binding cassette, sub-family F (GCN20), member 1	BC034488.2	8117929	BC034488.2:420
ARF1/MIR3620	ADP-ribosylation factor 1/microRNA 3620	NM_001658.3	7910229	NM_001658.3:1446
AVL9	AVL9 homolog	NM_015060.1	8132188	NM_015060.1:418
B2M	beta-2-microglobulin	NM_004048.2	7983360	NM_004048.2:573
CLTC	clathrin, heavy chain (Hc)	NM_004859.3	8008834	NM_004859.3:6144
CSNK1D	casein kinase 1, delta	NM_001893.3	8019463	NM_001893.3:1535
G6PD	Glucose-6-phosphate dehydrogenase	NM_001042351.2	8176133	NM_001042351.2:2020
GNAS	GNAS complex locus	NM_001077488	8063668	GPS_8063668.1:4962
GNB1	guanine nucleotide binding protein (G protein), polypeptide 1	NM_002074.2	7911619	NM_002074.2:2165
HPRT1	hypoxanthine phosphoribosyltransferase 1	NM_000194.1	8169984	NM_000194.1:240
LDHA	lactate dehydrogenase A	NM_005566.3	7938777	NM_005566.3:2123
LDHC	lactate dehydrogenase C	NM_017448.3	7938788	NM_017448.3:551
PGK1	phosphoglycerate kinase 1	NM_000291.3	8168500	NM_000291.3:192
POLR2A	polymerase (RNA) II (DNA directed) polypeptide A, 220kDa	NM_000937.2	8004431	NM_000937.2:3775
RPL19	ribosomal protein L19	NM_000981.3	8006845	NM_000981.3:582
RPLP0	ribosomal protein, large, P0	NM_053275.3	7966996	NM_053275.3:72
SDHA	succinate dehydrogenase complex, subunit A, flavoprotein	NM_004168.1	8104166	NM_004168.1:230
TBP	TATA box binding protein	NM_003194.4	8123507	NM_003194.4:602
TRIM27 (2)	tripartite motif containing 27	AJ420519.1	8124610	AJ420519.1:2234
TRIM27 (1)	tripartite motif containing 27	NM_006510.4	8179575	NM_006510.4:1724
UBA52	ubiquitin A-52 residue ribosomal protein fusion product 1	NM_001033930.1	8027024	NM_001033930.1:677
UBE2D2	Ubiquitin-conjugating enzyme E2D 2	NM_181838.1	8108435	NM_181838.1:698

Table S3. SVM classifications of compounds in independent test sets.

Treatment	SVM decision values (Calibrated)			Classification
	1	2	3	Pos if median > 0
Testset 1 - Sensitizers				
2-Nitro-1,4-phenylenediamine	4.66	3.93	5.22	Sensitizer
4-Nitrobenzyl bromide	0.28	1.44	1.54	Sensitizer
7-Hydroxycitronellal	1.44	4.17	2.43	Sensitizer
Benzalkonium chloride	0.86	0.37	1.82	Sensitizer
Diethyl maleate	9.06	8.92	7.85	Sensitizer
Diphenylcyclopropanone	1.16	5.06	4.13	Sensitizer
Ethyl acrylate	2.7	0.58	2.97	Sensitizer
Hydroquinone	3.03	7.06	5.02	Sensitizer
Lauryl gallate	-0.41	-1.43	-1.98	Non-Sensitizer ^a
Phenylacetaldehyde	-0.63	2.43	1.93	Sensitizer
Tetramethylthiuram disulfide	9.19	7.55	4.74	Sensitizer
Testset 1 – Non-sensitizers				
Dimethyl isophthalate	-2.7	-2.33	-0.96	Non-Sensitizer
Hexane	-1.09	-1.02	-3.8	Non-Sensitizer
Lactic acid	-3.29	-3.52	-2.17	Non-Sensitizer
Menthol	-1.35	-0.86	-1.58	Non-Sensitizer
Methyl salicylate	-0.75	-2.24	-0.2	Non-Sensitizer
Saccharin	-2.58	-1.04	-1.81	Non-Sensitizer
Salicylic acid	-1.87	-2.31	-2.33	Non-Sensitizer
Sodium dodecyl sulphate	0.41	1.64	1.95	Sensitizer ^a
Xylenes	-0.31	-0.52	-2.39	Non-Sensitizer
Testset 2 - Sensitizers				
2,4-Dinitrofluorobenzene	0.97	0.87	2.38	Sensitizer
3-Methylcatechol	4.04	4.49	3.71	Sensitizer
Butyl glycidyl ether	4.15	5.09	3.32	Sensitizer
Diethylthiocarbamyl benzothiazole sulfide	8.95	10.41	5.84	Sensitizer
Dimethylthiocarbamyl benzothiazole sulfide	14.07	15.51	12.88	Sensitizer
Lyril	0.09	0.04	1.74	Sensitizer
Testset 2 – Non-sensitizers				
1-Bromobutane	1.56	-1.6	2.53	Sensitizer ^a
Benzoic acid	-0.37	0.31	-0.34	Non-Sensitizer
Citric acid	0.24	-1.06	-0.16	Non-Sensitizer

^aIndicated compounds were falsely classified by the calibrated model.

Figure S1. Expression stability of candidate reference genes.



GeNorm and NormFinder algorithms were used for ranking of candidate reference genes (n=24). Values shown for the GeNorm algorithm represents the average expression stability values (M) of remaining reference genes when stepwise excluding the least stable candidate, and recalculating the M values for remaining genes. Values shown for NormFinder represents stability measures based on sample subgroup annotations (sensitizers and non-sensitizers), and are provided directly by the algorithm. Genes are sorted in order of expression stability, based on NormFinder stability values for each gene, with the most stable reference genes to the right (lowest stability values).