Research Article

The GARD Platform for Potency Assessment of Skin Sensitizing Chemicals

Kathrin S. Zeller [1], Andy Forreryd [1], Tim Lindberg [1], Robin Gradin [1,2], Aakash Chawade [3] and Malin Lindstedt [1]

[1]Department of Immunotechnology, Medicon Village, Lund, Sweden; [2]SenzaGen AB, Medicon Village, Lund, Sweden; [3]Swedish University of Agricultural Sciences, Alnarp, Sweden

Contact allergy induced by certain chemicals is a common health concern, and several alternative methods have been developed to fulfill the requirements of European legislation with regard to hazard assessment of potential skin sensitizers. However, validated methods, which provide information about the potency of skin sensitizers, are still lacking. The cell-based assay Genomic Allergen Rapid Detection (GARD), targeting key event 3, dendritic cell activation, of the skin sensitization AOP, uses gene expression profiling and a machine learning approach for the prediction of chemicals as sensitizers or non-sensitizers. Based on the GARD platform, we here expanded the assay to predict three sensitizer potency classes according to the European Classification, Labelling and Packaging (CLP) Regulation, targeting categories 1A (strong), 1B (weak) and no cat (non-sensitizer). Using a random forest approach and 70 training samples, a potential biomarker signature of 52 transcripts was identified. The resulting model could predict an independent test set consisting of 18 chemicals, six from each CLP category and all previously unseen to the model, with an overall accuracy of 78%. Importantly, the model was shown to be conservative and only underestimated the class label of one chemical. Furthermore, an association of defined chemical protein reactivity with distinct biological pathways illustrates that our transcriptional approach can reveal information contributing to the understanding of underlying mechanisms in sensitization.

Keywords: *in vitro* assay, sensitization, potency, biomarkers, random forest

1 Introduction

Skin sensitization and associated diseases such as contact allergy affect a substantial portion of the general population with an estimated prevalence of 15-20% in industrialized countries (Peiser et al., 2012). Allergic contact dermatitis (ACD), a type IV hypersensitivity reaction, is common among certain occupational groups such as those regularly exposed to chemicals or involved in wet work (Behroozy and Keegel, 2014). However, also cosmetics and household products can contain numerous skin sensitizing chemicals. The European Union has imposed requirements for testing of > 60,000 chemicals in the context of REACH (Hartung and Rovida, 2009), and prohibited animal testing for cosmetics and their ingredients (European Parliament, 2009). Information on both the skin sensitizing capacity and the potency of a chemical has to be provided to meet the regulatory requirements for classification and sub-categorization. Animal-free alternative assays that meet these requirements are urgently needed.

The molecular events leading to skin sensitization and consequently to ACD can be characterized by a number of sequential key events (KE) triggered by a chemical, and have been summarized in an adverse outcome pathway (AOP) as described by the OECD (2012). The initiating event (KE1) is defined as covalent protein modification by the skin sensitizing chemical after it has gained access to deeper skin layers. The following KE2 represents the inflammatory responses upon activation of keratinocytes. KE3 corresponds to the activation of dendritic cells, which in turn leads to activation and proliferation of T cells (KE4). Upon re-exposure to the sensitizer, the development of ACD may be triggered, which is characterized by skin lesions induced by specific Th1 and $CD8^+$ T cells. While the KE in the AOP are well described, a detailed mechanistic understanding of the underlying biology of the individual key events is still missing (OECD, 2012).

The murine Local Lymph Node Assay (LLNA) (Gerberick et al., 2007), followed by the Guinea Pig Maximization Test (Magnusson and Kligman, 1969), has for many years been the

Received January 10, 2017;
Accepted March 28, 2017;
Epub April 12, 2017;
doi:10.14573/altex.1701101

preferred assay for skin sensitization testing as the LLNA is able to provide data for both hazard identification and characterization, including skin sensitizer potency information. However, it is characterized by certain limitations such as susceptibility to vehicle effects and issues with false-positive results (Anderson et al., 2011). Several non-animal predictive methods have been developed to reduce animal experimentation used for chemical testing including computational approaches to integrate data from different test platforms for hazard identification as recently reviewed (Ezendam et al., 2016). Three test methods for skin sensitization are accepted as test guidelines at the OECD; the ARE-NRF2 luciferase method (KeratinoSens™) assay (Andreas et al., 2011; Natsch and Emter, 2008), the Direct Peptide Reactivity Assay (DPRA) (Gerberick et al., 2004) and the human Cell Line Activation Test (h-CLAT) (Ashikaga et al., 2006). In addition to hazard identification, information on skin sensitizer potency is imperative in order to allow quantitative risk assessment and to define exposure limits. Approaches for the prediction of skin sensitizer potency have been published and were recently reviewed by Ezendam et al. (2016), such as assays targeting KE2 (epidermal equivalent sensitizer potency assay (Teunis et al., 2014), SENS-IS (Cottrez et al., 2015)) and the U-SENS assay modelling KE3 (Piroird et al., 2015). Furthermore, *in silico* models, often combining information from several *in vitro* methods, have been described, for example QSAR (Dearden et al., 2015), artificial neural networks (Tsujita-Inoue et al., 2014), probabilistic models and integrated testing strategy (ITS) approaches including a Bayesian model (Jaworska et al., 2013, 2015; Luechtefeld et al., 2015; Natsch et al., 2015).

The alternative assay Genomic Allergen Rapid Detection (GARD) for the binary classification of chemicals into skin sensitizers and non-sensitizers is based on global transcriptomic analysis of differential expression in a human myeloid cell line, induced by sensitizing chemicals in comparison to non-sensitizing controls. The resulting biomarker signature, the GARD prediction signature (GPS), consists of 200 transcripts, which are used as input into a support vector machine (SVM) model trained on a set of reference chemicals (Johansson et al., 2011). The changes in transcription can be linked to the maturation and activation of dendritic cells (KE3) during sensitization. In an in-house study based on 26 blinded chemicals, the accuracy of the assay was estimated to be 89% (Johansson et al., 2014) primarily based on LLNA reference data.

Previous observations indicated that the GARD assay is able to provide information relevant also for potency assessment. Firstly, signaling pathways were differentially regulated depending on the potency of a subset of chemical reactivity groups (Albrekt et al., 2014). Secondly, we observed that more potent sensitizers were generally assigned higher GARD SVM decision values compared to weaker sensitizers, indicating that there were genes within the signature contributing with potency information (unpublished observations). However, the information in the GARD prediction signature was not sufficient to completely stratify chemicals into the well-defined potency groups as described by the Classification, Labelling and Packaging (CLP) Regulation (CLP, 2016).

The CLP Regulation is based on the Globally Harmonised System (CLP, 2016), and uses three categories for chemical classification: no category (no cat) for non-sensitizers, category 1B for weak and 1A for strong sensitizers. In light of the above described observations, it was hypothesized that GARD can be developed further into a tool for the prediction of chemical skin sensitizer potency, targeting the CLP categories. As the established GARD SVM model cannot be applied to multiclass problems, we used another approach based on random forest modelling (Breiman, 2001). Random forest is a decision-tree based method and well-suited for microarray data (Díaz-Uriarte and Alvarez de Andrés, 2006). It divides the dataset internally and repeatedly into a training and test set through random sub-sampling (bootstrapping). Samples in the test set, referred to as out-of-bag samples, comprise approximately one third of the entire dataset, and are used in order to estimate the out-of-bag (OOB) error, i.e., the classification error.

Here, we present a new approach to predict skin sensitizer potency according to CLP categories based on supervised machine learning using a random forest model. Firstly, the global gene expression data from a training set comprising 68 unique chemicals and 2 vehicle control samples were used as input into a random forest model. The random forest model was subsequently combined with an algorithm for backward variable elimination. The algorithm initially ranked the variable importance of each transcript from the microarrays, and then iteratively fitted new random forests, while removing the least important variables from the previous iteration. Using this strategy, we were able to identify a set of 52 transcripts with the smallest OOB error rate when predicting the out-of-bag samples from the training set. The predictive performance of the 52 transcripts was challenged with an independent test set containing 18 chemicals previously unseen to the model. The chemicals in this test set could be predicted with an overall accuracy of 78%. In addition to the predictive model, we also demonstrated the versatility of analyzing whole transcriptomes of cells by performing pathway analysis to further improve the mechanistic understanding of skin sensitizing potency on a cellular level, confirming the hypothesis that different chemical reactivity classes induce distinct signaling pathways.

2 Materials and methods

Cells and flow cytometry

The myeloid cell line used in this study was derived from MUTZ-3 (DSMZ, Braunschweig, Germany) and maintained as described (Johansson et al., 2011, 2013). A phenotypic control analysis of the cells prior to each experiment was carried out by flow cytometry in order to confirm the cells' immature state. The following monoclonal antibodies were used: CD1a (DakoCytomation, Glostrup, Denmark), CD34, CD86, HLA-DR (BD Biosciences, San Jose, USA), all FITC-conjugated; CD14 (DakoCytomation), CD54, CD80 (BD Biosciences), all PE-conjugated. FITC- and PE-conjugated mouse IgG1 (BD Biosciences) served as isotype controls and propidium iodide as a marker for non-viable cells (BD Biosciences). Three batches

Tab. 1: The 37 novel chemicals with CLP annotations used to complement existing GARD data

Name	CAS#	CLP	Cytotox	GARD input (M)	Binary class (HP*)	GARD binary prediction
2,4-dinitrofluorobenzene	70-34-8	1A	yes	0.00001	S	S
3-methylcatechol	488-17-5	1A	yes	0.00004	S	S
bisphenol A-diglycidyl ether	1675-54-3	1A	yes	0.00005	S	S
chlorpromazine	50-53-3	1A	yes	0.0000125	S	S
cyanuric chloride	108-77-0	1A	yes	0.00005	S	NS
glutaraldehyde	111-30-8	1A	yes	0.00002	S	S
hexyl salicylate	6259-76-3	1A	yes	0.00007	S	S
iodopropynyl butylcarbamate	55406-53-6	1A	yes	0.00001	S	S
methyl heptine carbonate	111-12-6	1A	yes	0.0001	S	S
p-benzoquinone	106-51-4	1A	yes	0.00005	S	S
propyl gallate	121-79-9	1A	yes	0.000125	S	S
abietic acid	514-10-3	1B	no	0.000125	S	S
amylcinnamyl alcohol	101-85-9	1B	yes	0.0003	S	S
anethole	104-46-1	1B	no	0.0005	NS	S
aniline	62-53-3	1B	no	0.0005	S	NS
anisyl alcohol	105-13-5	1B	no	0.0005	NS	NS
benzocaine	94-09-7	1B	no	0.0005	S	NS
benzyl benzoate	120-51-4	1B	yes	0.0003	NS	S
butyl glycidyl ether	2426-08-6	1B	yes	0.0005	S	NS
citral	5392-40-5	1B	yes	0.0000625	S	S
citronellol	106-22-9	1B	no	0.0005	NS	S
diethanolamine	111-42-2	1B	no	0.0005	NS	NS
imidazolidinyl urea	39236-46-9	1B	yes	0.00005	S	S
isopropyl myristate	110-27-0	1B	no	0.0005	NS	NS
lilial	80-54-6	1B	yes	0.0001875	S	S
limonene	5989-27-5	1B	no	0.0005	S	NS
linalool	78-70-6	1B	no	0.0005	S	NS
lyral	31906-04-4	1B	yes	0.0001	S	S
pentachlorophenol	87-86-5	1B	no	0.0000625	NS	NS
pyridine	110-86-1	1B	no	0.0005	NS	NS
1-bromobutane	109-65-9	no cat	no	0.0005	NS	NS
benzoic acid	65-85-0	no cat	no	0.0005	NS	NS
benzyl alcohol	100-51-6	no cat	no	0.0005	NS	NS
citric acid	77-92-9	no cat	no	0.0005	NS	NS
dextran	9004-54-0	no cat	no	0.00003	NS	NS
kanamycin A	25389-94-0	no cat	no	0.000125	NS	NS
tartaric acid	87-69-4	no cat	no	0.0005	NS	NS

* NS, non-sensitizer; S, sensitizer; based on Basketter et al. (2014) where available; HP, human potency (1-4 = S; 5-6 = NS). Otherwise according to CLP, see also Table 4.

of cells, representing three biological replicates, were exposed for 24 h in independent experiments and viability and CD86 expression were assessed by flow cytometry. All FACS samples were analyzed on a FACSCanto II instrument with FACS Diva software for data acquisition. 10,000 events were acquired and further analysis was performed in FCS Express V4 (De Novo Software, Los Angeles, CA). Cells for RNA extraction were lysed in TRIzol® (Life Technologies/Thermo Fisher Scientific, Waltham, USA) and stored until further use at -20°C.

Chemicals and stimulations

All chemicals were purchased from Sigma Aldrich (Saint Louis, USA) in high purity quality or they were provided by Cosmetics Europe. All chemicals were stored according to the recommendations of the supplier. The chemical stimulation of cells was performed as described earlier (Johansson et al., 2013). In short, GARD input concentrations were defined by solubility and cytotoxicity characteristics of the chemicals. An end concentration of 500 µM was targeted for non-cytotoxic and soluble chemicals and the highest possible concentration for chemicals with limited solubility (lower than 500 µM in medium). Cytotoxic chemicals were used at a concentration targeting a relative viability of cells of 90%. Most chemicals were used from a 1,000x pre-dilution in dimethyl sulfoxide (DMSO) or autoclaved MilliQ water. DMSO concentration as vehicle never exceeded 0.1%. DMSO and MilliQ samples were included as vehicle controls in this study and thus belong to the group of non-sensitizer samples.

RNA extraction, cDNA and array hybridization

RNA isolation from cells lysed in TRIzol® was performed according to the manufacturer's instructions. All samples were subjected to quality control using a Bioanalyzer 2100 (Agilent, Santa Clara, United States) prior to hybridization to the microarrays. Labeled sense DNA was synthesized according to Affymetrix (Affymetrix, Cleveland, USA) protocols using the recommended kits and controls. The cDNA was hybridized to Human Gene 1.0 ST arrays (Affymetrix) and further processed and scanned as recommended by the supplier. The new microarray data were merged with historical data (Johansson et al., 2011, 2014) and together subjected to quality control. The low-quality arrays excluded from downstream processing were identified from the normalized unscaled standard error (NUSE), which is an established measure to estimate overall variation for a specific array. An array was defined as poor quality if it was centered around 1.1 or had an overall higher spread of the NUSE distribution than the others, according to recommendations provided by Affymetrix. Additionally, the poor quality was also confirmed from the distribution of the arrays in a PCA plot.

Chemical classifications and dataset overview

A novel dataset comprising 37 well-characterized chemicals (Tab. 1) was selected in order to complement historical GARD data for 49 chemicals. The novel chemicals were selected based on European Chemicals Agency CLP databases and literature (Basketter et al., 2014; Piroird et al., 2015) and in cooperation with the Skin Tolerance Task Force of Cosmetics Europe (CE), who kindly provided 27 chemicals. In order to build a random forest model, the new microarray data were merged with historical data (Johansson et al., 2011, 2014), resulting in information from 86 unique chemicals and two vehicle controls (Tab. 2). Thereof, 68 samples were defined as a training set, and 18 samples, corresponding to six chemicals from each CLP category, were included in the independent test set. For four chemicals, CLP classification 1B was changed to no category / non-sensitizer according to the sources indicated in Table 2 for one of three reasons: i) for retaining consistency with previous GARD projects (benzaldehyde, xylene), ii) for being used as vehicle at a non-sensitizing concentration (DMSO), and for being a well-described false-positive in the LLNA (sodium dodecyl sulfate).

Binary classifications

Binary classifications of the 37 chemicals summarized in Table 1 into sensitizers or non-sensitizers were performed with the previously established model based on SVM, using SCAN-normalized (Piccolo et al., 2012, 2013) expression data from the GPS as variable input into the learning algorithm (Johansson et al., 2011). Prior to model construction, potential batch effects between training set and test chemicals were eliminated by scaling array expression values for test chemicals against the training set. A scaling factor was generated by calculating the ratio of the average expression value for each transcript in DMSO vehicle control samples of the training set and the average expression value for same transcript in DMSO samples in the batch where the test chemical originated. The scaling factor for each transcript was then multiplied with the expression values for the corresponding transcript in the test chemical. SVM predictions were performed as described previously (Forreryd et al., 2016; Johansson et al., 2011). In short, an SVM model based on a linear kernel was trained on reference chemical stimulations from the original training set used during identification of the GPS (Johansson et al., 2011). The trained model was subsequently applied to assign each test chemical with an SVM decision value (SVM DV). Resulting SVM DV for all test chemicals were used to construct a receiver operating characteristics (ROC) curve, and the resulting area under the curve (AUC) was used as a classification measure (Lasko et al., 2005). SVM modeling and ROC curve visualizations were performed in R statistical environment, using the additional packages e1071 (Dimitriadou, 2011) and ROCR (Sing et al., 2005). Prior to evaluating final predictive performance of the model, SVM DVs for each individual replicate of the test chemicals were calibrated against the cut-off for maximal predictive performance obtained during classification of benchmark samples in Table 3, as described by Forreryd et al. (2016). The calibrated SVM DVs were subsequently used for final classifications, and test chemicals were classified as sensitizers if the median output value of replicates > 0. Accuracy, sensitivity and specificity was estimated using Cooper statistics (Cooper et al., 1979). The non-parametric two-sample Wilcoxon test was performed in order to determine if the SVM DV distributions between CLP categories 1A, 1B and no cat differed significantly.

Tab. 2: Controls and 86 unique chemicals used to train and test the random forest model for the prediction of CLP categories

Stimulation	HP	CLP	Binary class	Set	Toxtree protein binding class
1-brombutane	na	no cat	NS	test	SN2
anethole	5	1B	S	test	MA
benzoic acid	na	no cat	NS	test	no binding
benzyl benzoate	5	1B	S	test	AT
bisphenol A-diglycidyl ether	3	1A	S	test	SN2
butyl glycidyl ether	3	1B	S	test	SN2
citric acid	na	no cat	NS	test	no binding
cyanuric chloride	na	1A	S	test	SNAr
diethyl maleate	2	1B	S	test	MA
diethyl phthalate	6	no cat	NS	test	no binding
ethyl vanillin	nf	no cat	NS	test	SB
glutaraldehyde	2	1A	S	test	SB
iodopropynyl butylcarbamate	4	1A	S	test	AT
linalool	4	1B	S	test	no binding
lyral	2	1B	S	test	SB
p-benzochinone	na	1A	S	test	MA
propyl gallate	2	1A	S	test	MA
xylene[1]	6	no cat	NS	test	no binding
1-butanol	6	no cat	NS	train	no binding
2,4-dinitrochlorobenzene	1	1A	S	train	SNAr
2,4-dinitrofluorobenzene	na	1A	S	train	SNAr
2-aminophenol	2	1A	S	train	MA
2-hydroxyethyl acrylate	3	1A	S	train	MA
2-mercaptobenzothiazole	3	1A	S	train	AT
2-nitro-1,4-phenylenediamine	2	1A	S	train	MA
3-methylcatechol	na	1A	S	train	MA
4-methylaminophenol sulfate	3	1A	S	train	MA
4-nitrobenzylbromide	na	1A	S	train	SN2
abietic acid	3	1B	S	train	no binding
amylcinnamyl alcohol	4	1B	S	train	MA
aniline	4	1B	S	train	no binding
anisyl alcohol	5	1B	S	train	MA/SN2
benzaldehyde[2]	5	no cat	NS	train	SB
benzocaine	4	1B	S	train	no binding
benzyl alcohol	na	no cat	NS	train	no binding
chloroanilin	na	1B	S	train	no binding
chlorobenzene	na	no cat	NS	train	no binding
chlorpromazine	3	1A	S	train	SB
cinnamaldehyde	2	1A	S	train	MA
cinnamyl alcohol	3	1B	S	train	MA
citral	3	1B	S	train	SB
citronellol	5	1B	S	train	no binding
dextran	6	no cat	NS	train	SB

diethanolamine	5	1B	S	train	no binding
dimethyl formamide	nf	no cat	NS	train	nf
dimethyl sulfoxide[3]	6	no cat	NS	train	no binding
diphenylcyclopropenone	1	1A	S	train	MA
ethylenediamine	3	1B	S	train	SB
eugenol	3	1B	S	train	MA
formaldehyde	2	1A	S	train	SB
geraniol	4	1B	S	train	SB
glycerol	6	no cat	NS	train	no binding
glyoxal	2	1A	S	train	no binding
hexane	6	no cat	NS	train	no binding
hexyl salicylate	4	1A	S	train	no binding
hexylcinnamic aldehyde	5	1B	S	train	MA
hydroquinone	3	1A	S	train	MA
hydroxycitronellal	4	1B	S	train	SB
imidazolidinyl urea	3	1B	S	train	AT
isoeugenol	2	1A	S	train	MA
isopropanol	5	no cat	NS	train	no binding
isopropyl myristate	5	1B	S	train	no binding
kanamycin A	6	no cat	NS	train	no binding
Kathon CG	1	1A	S	train	nf
lactic acid	6	no cat	NS	train	no binding
lauryl gallate	2	1A	S	train	MA
lilial	4	1B	S	train	SB
limonene	4	1B	S	train	no binding
methyl heptine carbonate	2	1A	S	train	MA
methyl salicylate	5	no cat	NS	train	no binding
methyldibromo glutaronitrile	2	1A	S	train	MA/SN2
octanoic acid	6	no cat	NS	train	no binding
pentachlorophenol	5	1B	S	train	SNAr
phenol	6	no cat	NS	train	no binding
phenyl benzoate	3	1B	S	train	AT
phenylacetaldehyde	na	1B	S	train	SB
p-hydroxybenzoic acid	nf	no cat	NS	train	no binding
potassium dichromate	1	1A	S	train	no binding
potassium permanganate	nf	no cat	NS	train	nf
p-phenylenediamine	1	1A	S	train	MA
pyridine	5	1B	S	train	no binding
resorcinol	4	1B	S	train	MA
salicylic acid	6	no cat	NS	train	no binding
sodium dodecyl sulfate[4]	6	no cat	NS	train	SN2
tartaric acid	na	no cat	NS	train	no binding
tetramethylthiuram disulfide	3	1B	S	train	no binding
Tween 80	6	no cat	NS	train	na
unstimulated	6	no cat	NS	train	nf

AT, acyl transfer agent; HP, human potency; MA, Michael acceptor; na, not available; nf, not found; NS, non-sensitizer; S, sensitizer; SB, Schiff base formation; SN2, bi-molecular nucleophilic substitution; SNAr, nucleophilic aromatic substitution
[1,3,4] NS according to Basketter et al. (2014); [2] NS according to Sens-it-iv project (Roggen and Blaauboer, 2013)

Tab. 3: Benchmark chemicals

Chemical	CAS	CLP	Binary class	HP	GARD input (M)
2,4-dinitrochlorobenzene	97-00-7	1A	S	1	0.000004
p-phenylenediamine	106-50-3	1A	S	1	0.000075
2-hydroxethylacrylate	818-61-1	1A	S	3	0.0001
2-nitro-1,4-phenylenediamine	5307-14-2	1A	S	2	0.0003
2-aminophenol	95-55-6	1A	S	2	0.0001
resorcinol	108-46-3	1B	S	4	0.0005
geraniol	106-24-1	1B	S	4	0.0005
hexylcinnamic aldehyde	101-86-0	1B	S	5	0.000032
benzaldehyde*	100-52-7	no cat	NS	5	0.00025
chlorobenzene	108-90-7	no cat	NS	6	0.000098
1-butanol	71-36-3	no cat	NS	6	0.0005

*NS according to Basketter et al. (2014). HP, human potency; NS, non-sensitizer; S, sensitizer

Data handling and statistical analysis

In order to build a random forest model, 68 chemical and two vehicle control samples analyzed as described above with Human Gene 1.0 ST arrays (33,297 transcripts, partly named genes, more commonly referred to as variables) were defined as a training set, and 18 samples, six chemicals from each CLP category, were included in the independent test set. Samples in the test set were not included in the construction of the model. The aim was to obtain a balanced training set representing all three CLP categories (Tab. 4) and different chemical reactivity groups as listed in Table 2 ("Toxtree protein binding class"; Patlewicz et al., 2008; Piroird et al., 2015). Most of the chemicals in the test set (14 out of 18) originated from the latest experimental campaign (Tab. 1), comprising 37 chemicals previously not investigated using the GARD assay. In the training set, roughly one third of samples (23 out of 68) were from this latest dataset. The vehicle samples were part of all projects and are thus present with higher replicate numbers.

The new microarray data were merged with historical data (Johansson et al., 2011, 2014), and four arrays were removed due to poor quality; however, no chemical was present in less than biological duplicates, i.e., based on cells derived from at least two different batches. Array data was imported into the R statistical environment and normalized using the SCANfast algorithm (Piccolo et al., 2012, 2013). As several experimental campaigns needed to be combined, this dataset was normalized using the ComBat method (Leek et al., 2014; Johnson et al., 2007) in order to remove batch effects between samples. At this time, the samples in the training set were separated from the samples in the test set. To avoid overfitting, only samples in the designated training set were used during identification of the predictive biomarker signature and for fitting of parameters to the classifier, and samples in the test set were set aside to validate the performance of the identified signature and the specified classifier. The predictive biomarker signature was identified by feeding normalized and batch corrected transcript intensities from individual samples in the training set into a random forest model (Breiman, 2001) combined with a backward elimination procedure in the varSelRF package (Diaz-Uriarte, 2007) in R/Bioconductor version 3.1.2. The initial forest used for ranking of variable importance was grown to 2,000 trees and all other parameters were kept at the default settings. The package iteratively fits and evaluates random forest models, at each iteration dropping 20% of the least important variables. The best performing set of variables was selected based on OOB error rates from all fitted random forests as the smallest number of transcripts within one standard error from the minimal error solution (i.e., 1 s.e. rule). The variable selection procedure was validated by estimating the prediction error rate by the .632+ bootstrap method of the varSelRF package using 100 bootstrap samples, and the importance of individual transcripts in the biomarker signature was validated by the frequency of appearance in bootstrap samples (referred to as validation call frequencies (VCF)). The predictive performance of the identified biomarker signature was validated by building a new forest in the random forest package (Liaw and Wiener, 2002), based on previous parameters, using only the samples in the training set and the selected transcripts in the biomarker signature as variable input. The model was applied to assign each individual replicate sample in the test set to a CLP cat-

Tab. 4: Training and test set composition

	Total number	CLP 1A	CLP 1B	CLP no cat
Training set	70	23	25	22
Test set	18	6	6	6

egory, and the majority vote across the biological replicate stimulations for each chemical was accepted as the predicted category. Heatmaps and PCA plots were constructed in Qlucore Omics Explorer (Qlucore AB, Lund, Sweden) in order to visualize the RNA expression data.

Pathway analysis

Pathway analysis was performed with the Key Pathway Advisor (KPA) tool version 16.6 (KPA, 2016), which provides a pathway analysis workflow to investigate, e.g., gene expression data. It associates differentially expressed genes with both upstream and downstream processes in order to allow biological interpretation. The investigated dataset, consisting of the SCANfast- and ComBat-normalized expression values of the test set and the training set, in total 308 samples, was first variance-filtered in order to remove variables with consistently low variance until approximately a third was left (10,009 variables, $\sigma/\sigma_{max} = 0.1478$). A multigroup comparison (ANOVA) comparing samples belonging to CLP no cat, 1B, and 1A was then applied in order to identify transcripts that were differentially regulated. The most significant 883 transcripts (false discovery rate FDR = 10^{-9}; p = 8.53 x 10^{-11}) were used as input into KPA (Affymetrix Exon IDs and respective p-value, overconnectivity analysis). In order to identify pathways associated with protein reactivity, the same variance-filtered dataset was filtered based on two-group comparisons (t-tests, Toxtree binding class "no binding" non-sensitizers (81 samples) versus "Michael acceptor" sensitizers (MA, 63 samples), "no binding" versus "Schiff base formation" (SB, 29 samples), "no binding" versus combined "bi-molecular nucleophilic substitution/nucleophilic aromatic substitution" (SN, 25 samples). Lists with the 500 most significantly regulated variables from each comparison, together with p-value and fold change (causal reasoning analysis), were then entered into the KPA tool for each protein reactivity group. The lowest p-value was reached when comparing MA samples to "no binding" (FDR = 3.65 x 10^{-7}), followed by SB (FDR = 2.3 x 10^{-5}) and SN (FDR = 2.44 x 10^{-5}).

3 Results

3.1 Binary classifications of 37 chemicals

A novel dataset comprising 37 well-characterized chemicals (Tab. 1) was selected in order to complement historical GARD data for 51 chemicals and to represent a relevant choice of chemicals, balanced in terms of chemical reactivity class and use in consumer products. The 37 novel chemicals were predicted as sensitizers or non-sensitizers using the GPS and previously established protocols based on SVM classifications. The SVM model was applied to assign each individual replicate sample with a SVM DV. Prior to final classifications, SVM DVs from the 37 samples were first calibrated against 11 benchmark samples (Tab. 3) included in the same sample batch as the test chemicals. For the purpose of evaluating binary predictions, we here decided to prioritize human data (Basketter et al., 2014) when available, where classes one to four correspond to sensitizers and five and six to non-sensitizer, instead of CLP classifications. Model performance predicting the 37 chemicals is summarized by an AUC ROC of 0.88, indicating a good discriminatory ability, and as illustrated in Figure 1A. The sensitivity, specificity and accuracy based on Cooper statistics were estimated as 73%, 80% and 76%, respectively. In combination with previously published data (Johansson et al., 2014) the updated predictive accuracy of GARD for binary classification of skin sensitizers is estimated as 84% based on a dataset comprising a total of 74 chemicals. When the chemicals in Table 1 were grouped according to CLP classification, and the respective SVM DV values obtained during classification of the 37 chemicals were summarized in a boxplot as presented in Figure 1B, a potency gradient emerged, as the stronger sensitizers were assigned higher SVM DVs in comparison to the weaker sensitizers in category 1B and the non-sensitizers in no cat. According to non-parametric two-sample Wilcoxon tests comparing SVM DV sample distributions, these groups differed significantly (no cat vs 1B: p = 2.8^{-5}; 1B vs 1A: p = 2.8^{-6}, no cat vs 1A: p = 3.5^{-12}). Although the differ-

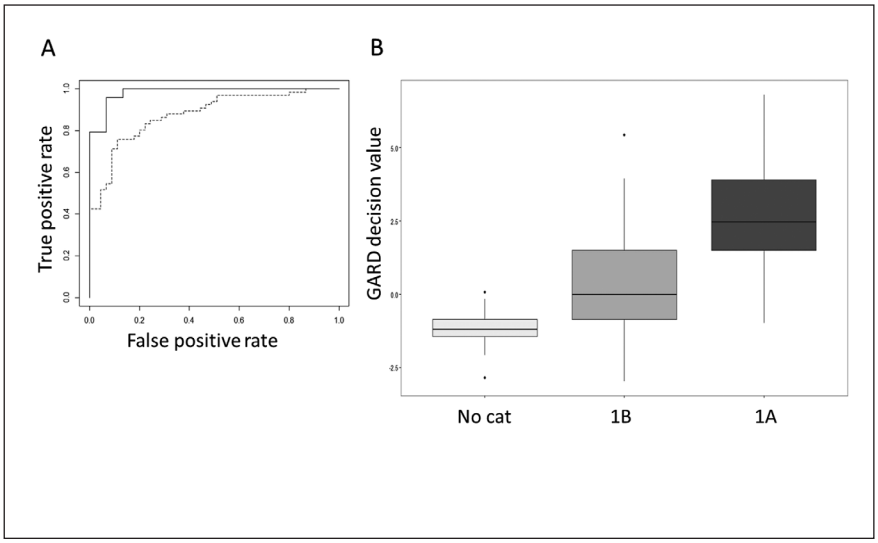

Fig. 1: Binary predictions using the GARD assay

ROC evaluation for (A) binary predictions of benchmark chemicals (filled line) and of 37 new chemicals (dotted line). (B) GARD SVM decision values (DVs) correlate with CLP potency (37 chemicals, 11 1A, 19 1B, 7 no cat). Increasing potency is associated with increasing DVs.

ences between groups were significant, some overlap existed between individual chemicals, indicating that the information was not sufficient to completely stratify samples into well-defined potency groups.

3.2 A random forest model for the prediction of CLP categories

In order to establish a biomarker signature for prediction of CLP categories, the dataset comprising the 37 novel chemicals was merged with the historical dataset comprising 51 chemicals. In total, the dataset consisted of 86 unique chemicals (Tab. 2) and two vehicle controls, balanced with regards to categories 1A and 1B and non-sensitizers (no cat) as described by CLP (Tab. 4). For four chemicals CLP classification 1B was changed to no category/non-sensitizer according to the sources indicated in Table 2 for one of three reasons: i) for retaining consistency with previous GARD projects (benzaldehyde, xylene), ii) for being used as vehicle in non-sensitizing concentration (DMSO), and for being a well-described false-positive in the LLNA (sodium dodecyl sulfate).

A random forest model for the prediction of three CLP categories was developed based on a training set consisting of 70 unique samples, including two vehicle controls. From an input of > 30,000 transcripts based on whole-genome array analysis, 52 predictive variables (transcripts) (Tab. 5) were identified as optimal for CLP classification. The model's prediction error rate, derived from bootstrapping, was estimated as 0.225, which provides an indication of model performance. In order to visualize the dataset used to develop the model, principal component analysis (PCA) was performed. The 52 variables identified by random forest were used as input, and the PCA was built on the training set (Fig. 2A). Figure 2A is based on chemicals with biological replicates colored according to CLP categories and a clear gradient from no cat to strong sensitizers (1A) can be observed along the first principal component. The heatmap of the training set with hierarchical clustering of the variables in Figure 2B illustrates the regulation of transcripts in relation to the respective chemical and CLP category.

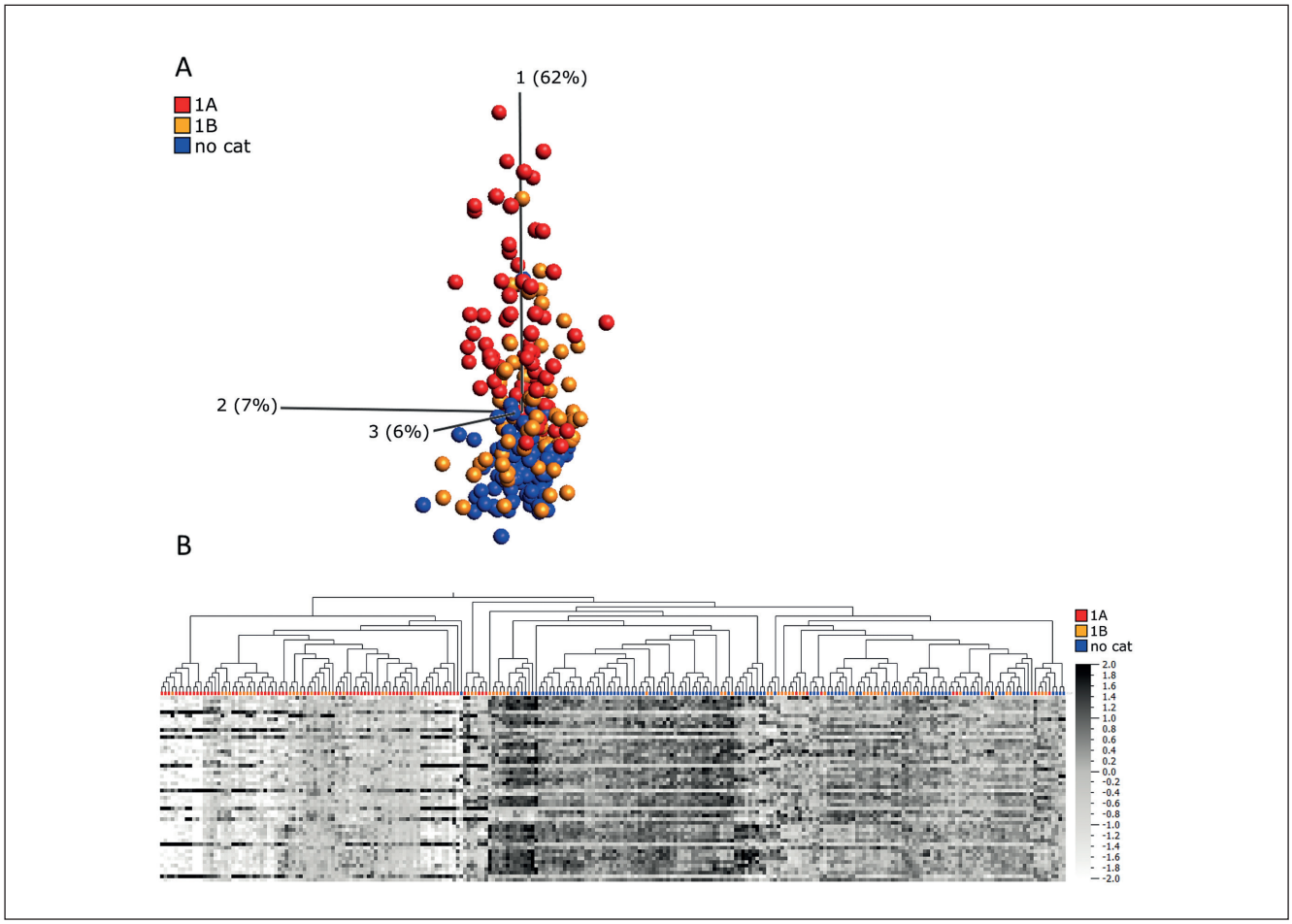

Fig. 2: Visualization of the training dataset used to develop the random forest model, using the 52 transcripts as input variables
(A) PCA plot of the training set (n = 70) with separate biological replicates (n = 254) colored according to CLP classifications of the chemicals. (B) Heatmap of the training set with replicates of the samples (x-axis) hierarchically clustered, where the grey scale represents the relative transcript expression intensity (y-axis).

Tab. 5: The 52 variables identified by random forest modeling as optimal for CLP classification
The variables marked in bold text are also found in the GARD GPS.

Transcript cluster ID	VCF (%)	Gene title	Gene symbol
8117594	93	histone cluster 1, H2bm	HIST1H2BM
8124385	86	histone cluster 1, H4b	HIST1H4B
8004804	**83**	**phosphoribosylformylglycinamidine synthase**	**PFAS**
8124430	81	histone cluster 1, H1d	HIST1H1D
8095221	80	phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase	PAICS
8124413	69	histone cluster 1, H4d	HIST1H4D
8005839	**63**	**transmembrane protein 97**	**TMEM97**
7916432	**61**	**24-dehydrocholesterol reductase**	**DHCR24**
8117608	56	histone cluster 1, H2al /// histone cluster 1, H2bn	HIST1H2AL /// HIST1H2BN
7994109	51	polo-like kinase 1	PLK1
7904433	44	phosphoglycerate dehydrogenase	PHGDH
8040843	44	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and anddihydroorotase	CAD
8082350	44	minichromosome maintenance complex component 2	MCM2
8141395	43	minichromosome maintenance complex component 7	MCM7
7898549	42	MRT4 homolog, ribosome maturation factor	MRTO4
7901091	41	target of EGR1, member 1 (nuclear)	TOE1
7903893	41	CD53 molecule	CD53
8118669	41	kinesin family member C1	KIFC1
7900699	40	cell division cycle 20	CDC20
7938348	40	WEE1 G2 checkpoint kinase	WEE1
8121087	36	peptidase M20 domain containing 2	PM20D2
8084630	35	NmrA-like family domain containing 1 pseudogene	LOC344887
7957737	34	thymopoietin	TMPO
8146357	34	minichromosome maintenance complex component 4	MCM4
7918300	33	proline/serine-rich coiled-coil 1	PSRC1
8054329	31	ring finger protein 149	RNF149
8055426	31	minichromosome maintenance complex component 6	MCM6
7948656	30	ferritin, heavy polypeptide 1	FTH1
7958455	30	uracil DNA glycosylase	UNG
8117408	**30**	**histone cluster 1, H2ae**	**HIST1H2AE**
8072687	29	minichromosome maintenance complex component 5	MCM5
8119088	27	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	CDKN1A
8117395	26	histone cluster 1, H2bf	HIST1H2BF
8124527	25	histone cluster 1, H1b	HIST1H1B
7896697	21	---	---
8003503	20	Fanconi anemia complementation group A	FANCA

8097417	20	jade family PHD finger 1	JADE1
7977445	18	KIAA0125	KIAA0125
7985213	17	cholinergic receptor, nicotinic alpha 5	CHRNA5
8002303	**17**	**NAD(P)H dehydrogenase, quinone 1**	**NQO1**
8068478	17	chromatin assembly factor 1, subunit B (p60) /// MORC family CW-type zinc finger 3	CHAF1B /// MORC3
8099721	16	sel-1 suppressor of lin-12-like 3 (*C. elegans*)	SEL1L3
7948192	14	structure specific recognition protein 1	SSRP1
7960340	14	forkhead box M1	FOXM1
8107706	14	lamin B1	LMNB1
8124524	14	histone cluster 1, H2ak	HIST1H2AK
8040712	11	centromere protein A	CENPA
8043602	10	non-SMC condensin I complex subunit H	NCAPH
7939341	8	CD44 molecule (Indian blood group)	CD44
8124394	7	histone cluster 1, H2bb	HIST1H2BB
8144931	7	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B2	ATP6V1B2
7999025	5	TNF receptor-associated protein 1	TRAP1

VCF, variable call frequency

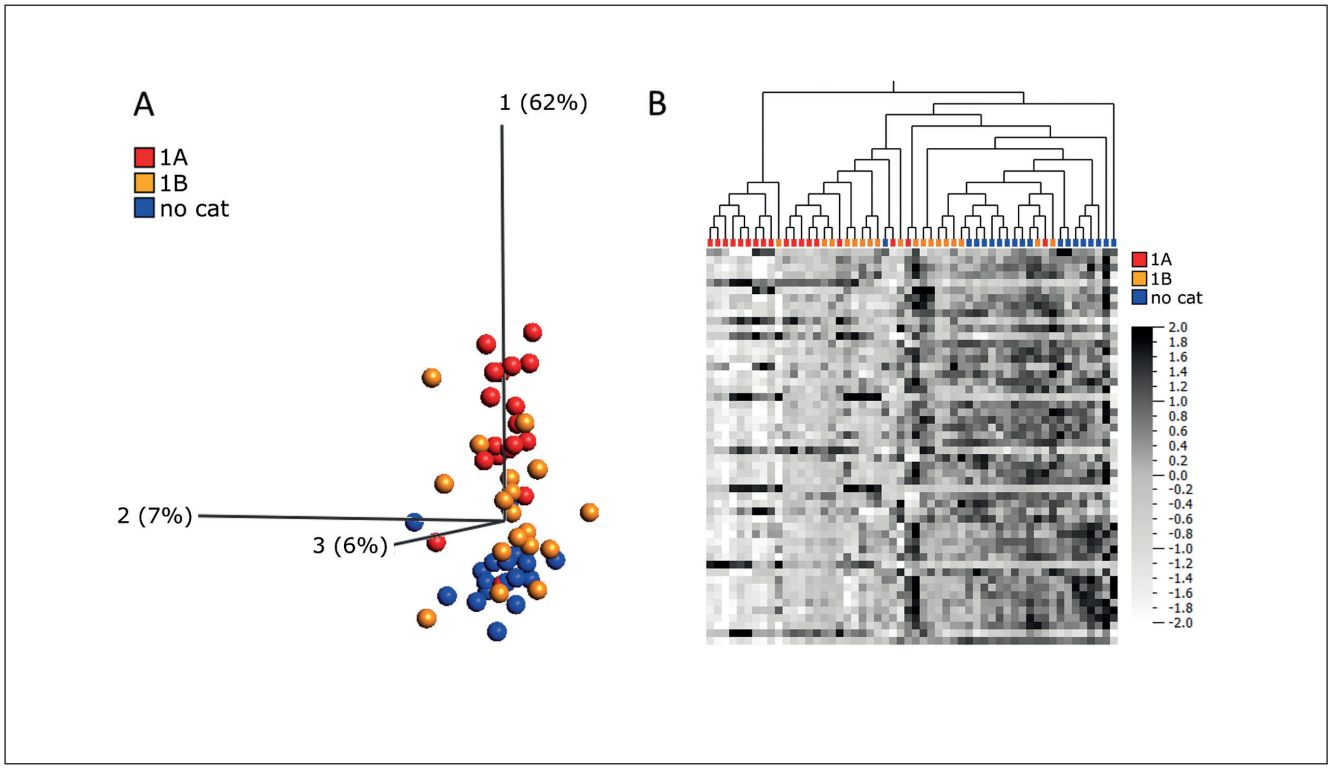

Fig. 3: Visualization of the test dataset (n = 18) using the 52 transcripts
(A) PCA plot of the test set with separate biological replicates (n = 3) colored according to CLP classifications. The PCA was built on the training set and the test set plotted without influencing the PCA. (B) Heatmap of the test set with samples (x-axis) hierarchically clustered, where the grey scale represents relative transcript expression intensity (y-axis).

3.3 Prediction of an independent test set

The model performance was further evaluated by predicting the CLP categories of an independent test set, which comprised 18 chemicals previously unseen by the model. The test set colored according to CLP categories was visualized in the PCA plot (Fig. 3A), without influencing the PCA components based on the 52 identified variables. Figure 3B visualizes the regulation of the 52 transcripts in the normalized test set in form of a heatmap. When replicates were predicted separately and majority voting was used to classify the respective chemicals, 14 out of 18 chemicals were assigned into the correct CLP category (Tab. 6; Tab. S1 at doi:10.14573/altex.1701101s), resulting in an overall accuracy of 78% (Tab. 7). The four misclassified chemicals were diethyl maleate, butyl glycidyl ether, lyral and cyanuric chloride. The only false-negative prediction was cyanuric chloride, which is classified as 1A in CLP and as no cat in our model, whereas the remaining three chemicals are classified as CLP 1B but were predicted as 1A. In a subsequent step to confirm that our selections of training and test set were unbiased and that the predictive model was not entirely dependent on the composition of the training set, we constructed 18 alternative random forest models, where the composition of chemicals in the training and test set were randomly shuffled. For each new model, we repeated the complete process of variable selection as described above. Dependent on the number of replicate samples available for each chemical stimulation, the total number of samples in each training and test set varied, but the number of chemicals in each set and their CLP distribution were kept constant. The alternative models were all significant and the

Tab. 7: Statistics by class for separate replicates in the test set predictions

Six chemicals per category; three biological replicates each; n (test set) = 18.

	no cat	1A	1B
Sensitivity	0.889	0.833	0.556
Specificity	0.917	0.806	0.917
Positive predictive value	0.842	0.682	0.769
Negative predictive value	0.943	0.902	0.805
Prevalence	0.333	0.333	0.333
Detection rate	0.296	0.278	0.185
Detection prevalence	0.352	0.407	0.241
Balanced accuracy	0.903	0.819	0.736

Tab. 6: Test set predictions using majority voting

Three biological replicates per chemical; for replicate predictions see Tab. S1. GARD misclassifications are highlighted in italics.

Chemical	true CLP	GARD predicted CLP	HP	Toxtree protein binding class
1-brombutane	no cat	no cat	na	SN2
benzoic acid	no cat	no cat	na	no binding
citric acid	no cat	no cat	na	no binding
diethyl phthalate	no cat	no cat	6	no binding
ethyl vanillin	no cat	no cat	nf	SB
xylene	no cat	no cat	6	no binding
anethole	1B	1B	5	MA
benzyl benzoate	1B	1B	5	AT
linalool	1B	1B	4	no binding
lyral	1B	***1A***	2	SB
butyl glycidyl ether	1B	***1A***	3	SN2
diethyl maleate	1B	***1A***	2	MA
cyanuric chloride	1A	***no cat***	na	SNAr
propyl gallate	1A	1A	2	MA
bisphenol A-diglycidyl ether	1A	1A	3	SN2
glutaraldehyde	1A	1A	2	SB
iodopropynyl butylcarbamate	1A	1A	4	AT
p-benzochinone	1A	1A	na	MA

AT, acyl transfer agent; HP, human potency; MA, Michael acceptor; na, not available; nf, not found; SB, Schiff base formation; SN2, bi-molecular nucleophilic substitution; SNAr, nucleophilic aromatic substitution

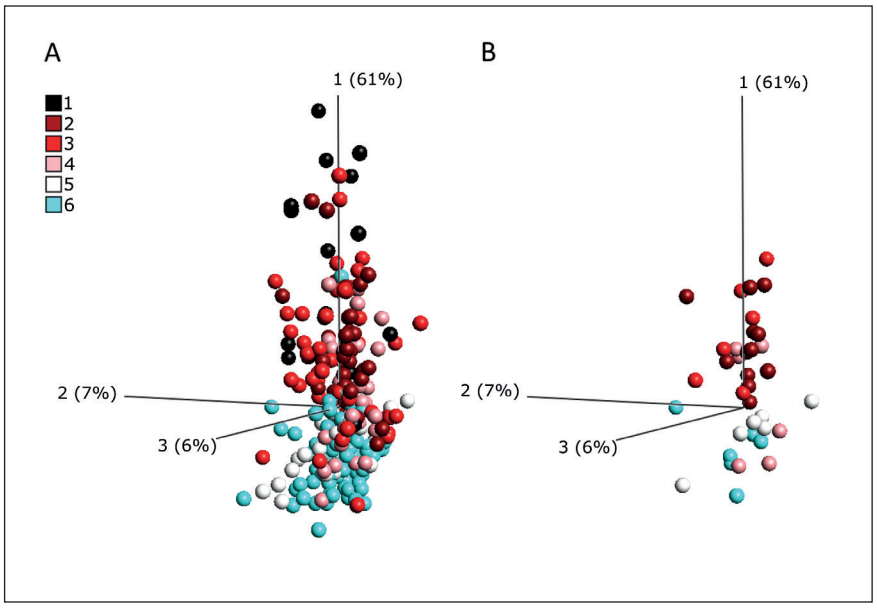

Fig. 4: The CLP potency model contains information related to human potency
(A) PCA plot of replicate training and test set samples with available human potency classifications (n = 71, 254 replicates) colored according to human potency. The PCA is based on the 52 random forest variables (transcripts) as input and the PCA was built on the training set. (B) PCA plot visualizing test set only (n = 12, 36 replicates).

average prediction error rate obtained from the bootstrapping procedure was identical to the initial model at 0.22, which supports that the presented model was not obtained due to a biased choice of training and test sets.

3.4 The CLP potency model contains information relevant for human potency prediction

Next, PCA was utilized in order to illustrate how the 52 variables (Tab. 5) perform using information related to human potency categories (Basketter et al., 2014). Notably, the PCA does not reflect the random forest model; it merely visualizes information contained in the 52 transcripts, and it is not possible to apply the CLP potency model directly to predict human potency classes. The 70 training and 18 test set samples were colored according to human potency (class 1-6, 6 = true non-sensitizer, 1 = strongest sensitizer), and samples for which no human potency category was available were removed (see Tab. 2). Although the model has been developed to predict CLP potency categories, the chosen 52 variables also contain information related to human potency as illustrated by the PCA visualization in Figure 4A, B.

3.5 Identity of the random forest model variables

The 52 variables identified by random forest (Tab. 5) represent transcripts belonging to different cell compartments and have different functional roles. Five of them overlap with the GARD prediction signature (GPS). Five of the top 10 markers, which were most frequently chosen in the bootstrapping process and have the highest validation call frequencies (Tab. 5), are histone cluster 1 members, such as HIST1H2AB (Singh et al., 2013). Histones are highly conserved and play an important role, not only for maintaining chromatin structure but also in gene regulation (Harshman et al., 2013). HIST1H2AE is both part of the GPS and the potency signature presented here. PFAS, also represented in the GPS, and PAICS are involved in purine biosynthesis (Lane and Fan, 2015; Li et al., 2007), and TMEM97 (part of GPS) is a regulator of cholesterol levels (Bartz et al., 2009), which is further described to be involved in cell cycle regulation, cell migration and invasion in a glioma cell model according to RNA interference experiments (Qiu et al., 2015). DHCR24 (another GPS transcript) represents a multifunctional enzyme localized to the endoplasmic reticulum (ER) that catalyzes the final step in cholesterol-synthesis (Waterham et al., 2001) but also possesses anti-apoptotic activity as, for example, shown for neuronal cells under ER stress (Lu et al., 2014). PLK1, a kinase, has been shown to be phosphorylated in response to TLR activation and results from RNA interference suggested that PLK1 signaling was involved in the TLR-induced inflammatory response (Hu et al., 2013). PLK1 was further reported to be involved in cell cycle regulation by inhibiting TNF-induced cyclin D1 expression and it could reduce TNF-induced NF-κB activation (Higashimoto et al., 2008). Many of the remaining transcripts are nuclear proteins and thus likely involved in DNA-dependent processes such as replication, transcription, splicing and cell cycle regulation. There are several further transcripts in the signature that code for proteins known for their involvement in immune responses and/or sensitization, such as NQO1, which is well-described for its role in the cellular response to skin sensitizers (Ade et al., 2009) and also part of the GPS. CD53 belongs to the tetraspanin family, transmembrane proteins that have multiple functions in, e.g,. cell adhesion, migration and signaling, which has been shown to be elevated on FcЄRI-positive skin DCs from atopic dermatitis patients (Peng et al., 2011) as well as on peripheral blood-derived monocytes from patients with atopic eczema (Jockers and Novak, 2006) in comparison to respective healthy controls. CD44 is a cell surface glycoprotein, adhesion and hyaluronan receptor (Lee-Sayer et al., 2015) expressed by numerous cell types and, for example, involved in inflammatory responses (Johnson and Ruffell, 2009), e.g.,

by mediating leukocyte migration into inflamed tissues, which has been shown in a mouse model of allergic dermatitis (Gonda et al., 2005).

3.6 Common and unique regulated pathways are induced by sensitizers differing in their protein reactivity

The 33 key pathways identified with an input of the 883 most significantly regulated transcripts after a multigroup comparison (FDR = 10^{-9}) between CLP categories in KPA analysis (Fig. 5) mirror several of the functional groups of the 52 variables defined by random forest, such as gene regulation, cell cycle control and metabolism. Immune response-associated pathways such as "IL-4-induced regulators of cell growth, survival, differentiation and metabolism" and "IL-3 signaling via JAK/STAT, p38, JNK and NF-κB" were among the 50% most significantly regulated ones.

The analyses subsequently focused on the three largest chemical reactivity groups in the present dataset; nucleophilic substitution (SN), Michael addition (MA) and Schiff base (SB) formation. Among the included chemicals, the majority of chemicals labeled as no cat possessed no protein binding properties; however, a few SB formation and SN chemicals were present. In category 1B, almost all protein reactivity types were represented, whereas there was a clear dominance of MA chemicals in category 1A.

For each associated protein reactivity, unique pathways could be identified for sensitizing chemicals belonging to the respective protein reactivity group as presented in Figure 6. These results combine differentially regulated transcripts from the input data with so-called key hubs, molecules that are able to regulate the expression level of the input transcripts (KPA, 2016). They cannot necessarily be identified themselves by gene expression experiments as their regulation may either be visible on other biological levels, such as activity changes (e.g., for kinases) or the changes may be very short-lived or of low magnitude. In total, 173 transcripts were common to all three reactivity groups (Fig. 7A) and six pathways were present in all three reactivity groups (Fig. 7B); "Cell cycle: Role of APC in cell cycle regulation", "Cell cycle: Role of SCF complex in cell cycle regulation", "Development: Transcription regulation of granulocyte development", "Cell cycle: Cell cycle (generic schema)", "DNA damage: ATM/ATR regulation of G1/S checkpoint", and "Mitogenic action of Estradiol / ESR1 (nuclear) in breast cancer". Again, cell cycle pathways were highly represented. Oxidative stress responses were identified as part of the key pathway results only for MA and SB chemicals (Fig. 6). In the MA sensitizer chemical group, KEAP1 and NRF2 were found as key hubs as well as their target genes NQO1 and HMOX1 (Ade et al., 2009; Natsch, 2010) and AHR (Schulz et al., 2013; Kohle and Bock, 2007). For MA chemicals, the target genes NQO1, HMOX1 and CES1 (Roberts et al., 2007) were even present on the input transcripts level. On the input level, CES1 was present for SN chemicals as well, but only NQO1, NRF2 and AHR were identified as key hubs. KEAP1 was not found as key hub in SB and SN KPA analysis and AHR was the only key hub identified for all three protein reactivity groups. NF-κB subunits (RelB or p52) were predicted key hubs in all reactivity groups except in MA chemicals.

In summary, there seem to be common mechanistic responses to chemical exposures *per se* such as cell cycle and DNA damage-related, but the pathway analysis results also support the hypothesis that different chemical reactivity classes induce distinct signaling pathways as observed earlier by us (Albrekt et al., 2014) and in other experimental systems (Cottrez et al., 2015; Migdal et al., 2013; Natsch et al., 2015; Chipinda et al., 2011). Several pathways are linked to processes known to be relevant in skin sensitization.

4 Discussion

The amount of chemical per exposed skin area that induces sensitization varies significantly (Basketter et al., 2014) among chemicals; thus, skin sensitizer potency information is imperative for accurate risk assessment. Developers of alternative test methods should rely on human clinical data in order to achieve high predictivity of human sensitization, however, this type of data is rather scarce and most available data is derived from the LLNA (Basketter et al., 2009). Despite the fact that animal models reflect the complexity of systemic diseases such as skin sensitization, *in vitro* data has so far shown to correlate well with and to perform equally well or even better than animal models, especially when combined in an ITS (Urbisch et al., 2015). Furthermore, alternative test systems may provide mechanistic insights that tests using whole animals cannot provide (Natsch et al., 2010).

Here, an approach to predict skin sensitizer potency is presented using the CLP system based on a dendritic cell (DC) model and transcriptional profiling. CLP categories are empirically determined and arbitrarily defined, which does not represent the diversity of different chemicals, their molecular features and mechanisms responsible for their sensitizing characteristics or the lack thereof. They are, however, what legislation currently requires in order to classify and label chemicals. We therefore investigated 37 additional chemicals previously not tested in the standard GARD assay in order to combine these new data with historical datasets. In the binary classifications of these new chemicals according to the established GARD model, four misclassified sensitizing chemicals were close to the cut-off as defined by the benchmark samples, namely aniline, benzocaine, limonene and butyl glycidyl ether. Three of these belong to human potency class 4, which shows that the model cut-off is critical in order to translate the SVM values, often correlating well with potency, into accurate classifications. Together with historical predictions, GARD shows an overall accuracy of 84% for binary classifications.

We then used both new and historical data in order to develop a random forest model for each CLP category, which displays balanced accuracies (Brodersen et al., 2010) of 96% for no category, 79% for category 1A and 75% for category 1B (based on majority votes for performance on

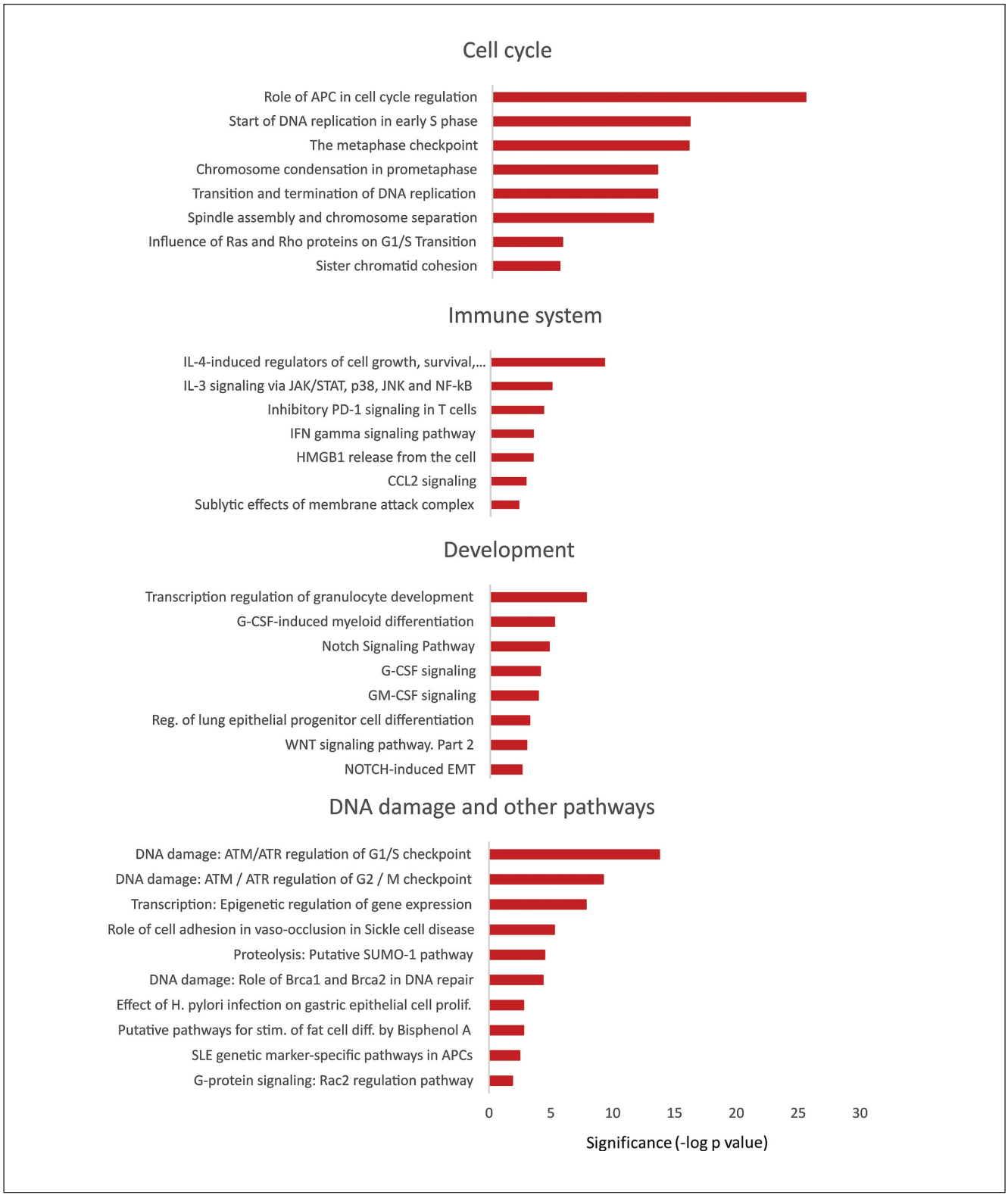

Fig. 5: Pathway analysis based on an input of the 883 most significant transcripts from a multigroup comparison of the three CLP classes using the Key Pathway Advisor tool

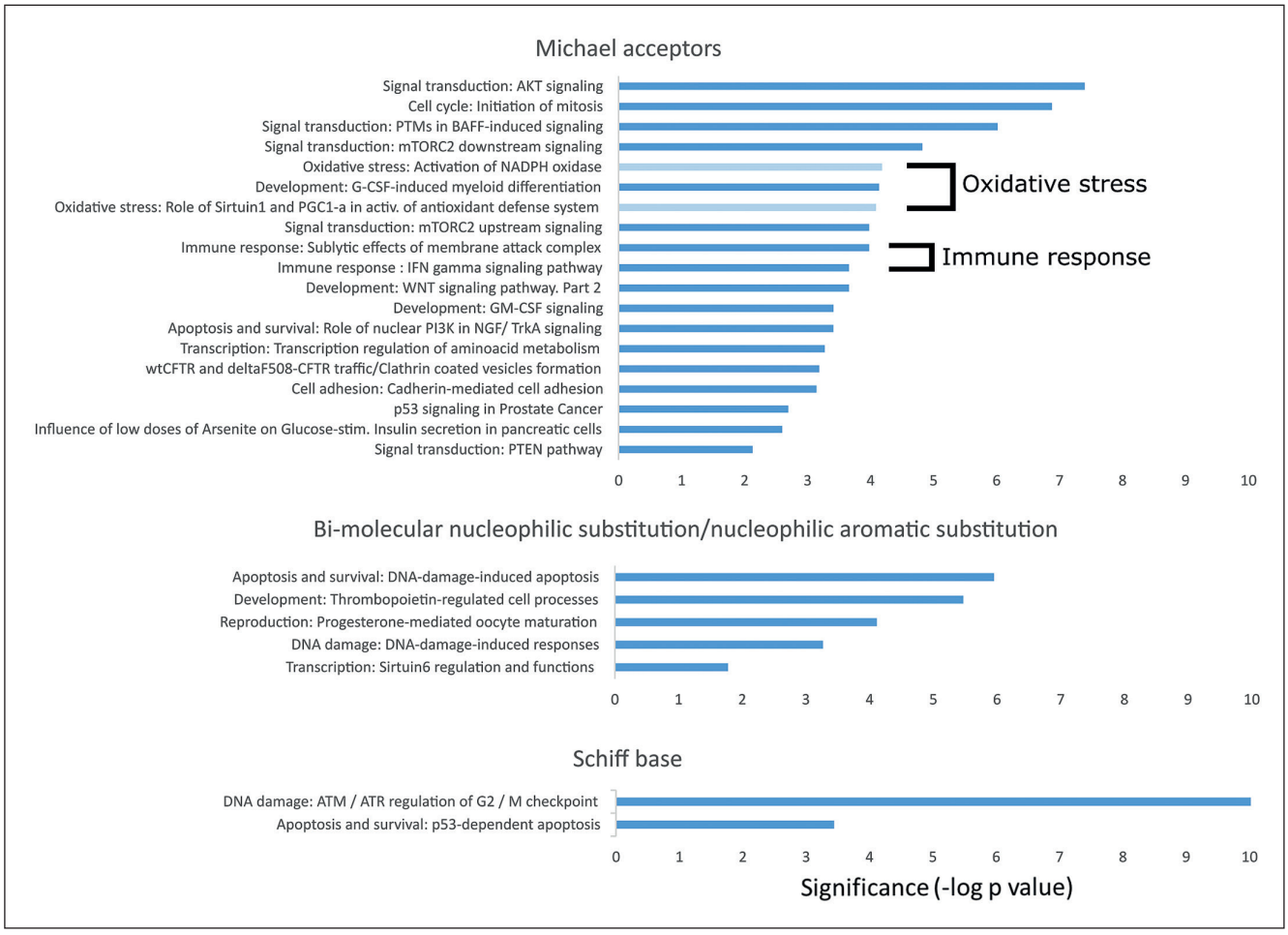

Fig. 6: Pathways unique for each protein reactivity group, as identified by pathway analysis using the Key Pathway Advisor tool

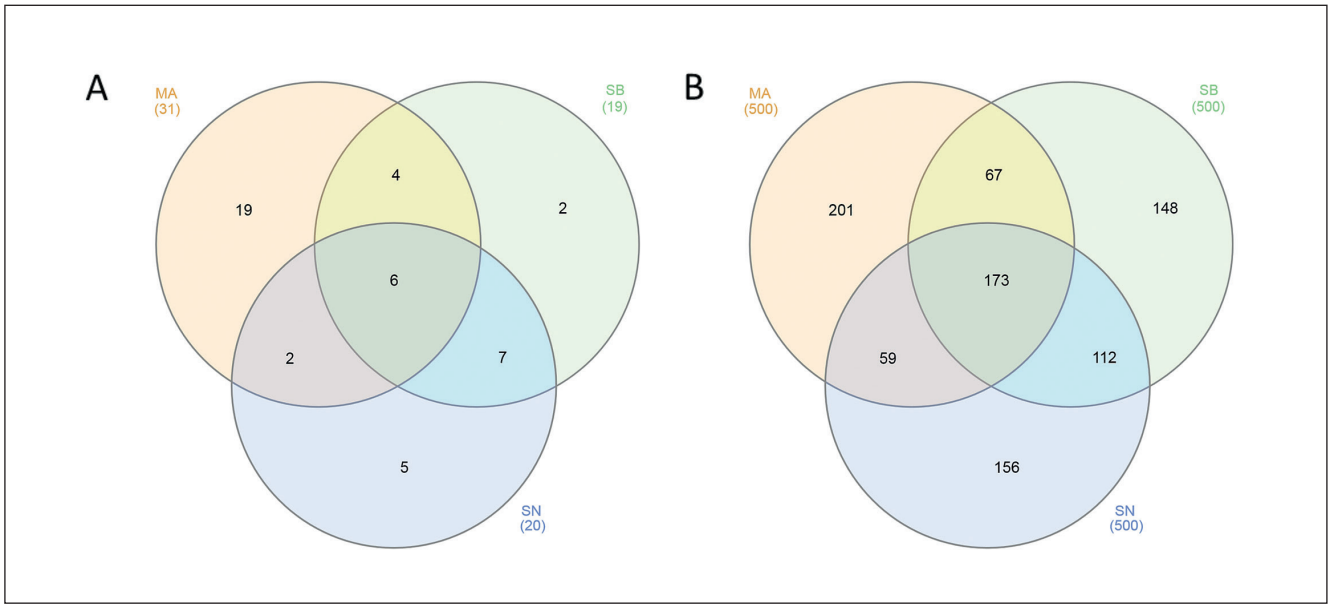

Fig. 7: Venn diagrams (Heberle et al., 2015) of common genes (A) and biological pathways (B) for the three different protein reactivity classes

MA = Michael acceptors; SB = Schiff base formation; SN = bi-molecular nucleophilic substitution/nucleophilic aromatic substitution.

replicate basis, see Tab. 7). Butyl glycidyl ether, diethyl maleate, cyanuric chloride and lyral were misclassified; the only false-negative prediction was no cat instead of 1A for cyanuric chloride. However, cyanuric chloride reacts exothermally with water, forming hydrogen chloride and possibly other reaction products. Due to this hydrolyzation reaction, probably already occurring in DMSO (containing water), the amount of cyanuric chloride and reaction products present in the assay are unknown. This chemical may fall outside the applicability domain of GARD platform-based assays. However, nothing in the quality control or other pre-modelling analyses motivated a removal of these samples. Diethyl maleate and lyral are classified as 1B by CLP, but as 1A by our model, which again seems to fit better with their human potency category, category 2, as described by Basketter et al. (2014). The forth misclassified chemical butyl glycidyl ether is a human potency category 3. Obviously, predicting 1B, i.e., weak sensitizers, seemed the most challenging part. Also in the LLNA, potency predictions of weak sensitizers vary more than those of strong sensitizers (Dumont et al., 2016; Hoffmann, 2015; Ezendam et al., 2016). Furthermore, 1B is a very heterogeneous group, both considering the range of LLNA EC3 concentrations (> 2%; CLP, 2016) and human potency categories associated to chemicals summarized in category 1B. In the dataset presented here, chemicals in category 1B belong to potency classes 2 to 5. Five of the 52 transcripts forming the potency prediction signature developed to distinguish three categories are also part of the GPS, which consists of in total 200 transcripts in order to predict two classes (Johansson et al., 2011). This degree of overlap may at first appear low, but considering a) the inherent differences in the development of the two models, i.e., the aim, in chemicals used as training set, in the applied bioinformatical methods, and b) the known "multiplicity problem" (Díaz-Uriarte and Alvarez de Andrés, 2006), it is not surprising. In short, the "multiplicity problem" describes the fact that variable selection with microarray data can result in several, equally good predictive models, in spite of sharing very few genes.

The U-SENS™ assay, formerly MUSST, uses another myeloid cell line, U937, and CD86 measurements in order to distinguish sensitizers and non-sensitizers. When the authors combined CD86 with cytotoxicity data and certain cut-off levels in order to predict CLP categories, correct predictions of 82% of Cat. 1A (41/50) and 73% of Cat. 1B/No Cat (85/116) were reported (Piroird et al., 2015). However, it remains unclear how the more challenging discrimination between no cat and 1B would turn out. Cottrez et al. (2016) recently published a study, where they report that their alternative assay SENSIS, a 3D reconstituted epidermis based model, performs very well for the prediction of skin sensitizer potency; however, they do not target CLP categories. Judging from Figure 4, purely based on the 52-variable input, which was defined in order to predict CLP categories, our model also seems to contain information relevant for human potency classification. Once more chemicals receive human potency classifications, it should be possible to smoothly develop a human potency model based on the GARD platform.

KPA pathway analysis identified biologically relevant events in the presented dataset as several pathways regulated have a known role in skin sensitization, e.g., cytokine signaling and oxidative stress responses (Fig. 5, 6). Although DCs are not the primary target for protein modification *in vivo*, we hypothesized that different protein reactivity classes influence the DC transcriptome differentially. Protein reactivity is one of the most important features of chemicals defining their skin sensitizing capacity and potency with certain limitations (Chipinda et al., 2011). Protein reactivity-specific patterns were detectable as revealed by the comparison of the most significantly regulated transcripts induced by the reactivity groups MA, SB, and SN. Interestingly, NF-κB subunits were predicted key hubs for all reactivity groups except for MA chemicals, which may reflect the described inhibitory effect on NF-κB signaling of this type of chemicals (Natsch et al., 2011). Although some pathways do not seem to fit into the context, such as "Mitogenic action of Estradiol / ESR1 (nuclear) in breast cancer", a closer look at regulated molecules reveals that those are certainly relevant also for other pathways. In this case, for example p21, c-myc, E2F1, SGOL2 (shugoshin 2) and CAD (carbamoyl phosphate synthetase) were involved, whereof the first ones are known cell cycle regulators/transcription factors (Buchmann et al., 1998) and play a role in chromosome segregation (SGOL2) (Xu et al., 2009). CAD, an enzyme, which is rate-limiting in the biosynthesis of pyrimidine nucleotides, on the other hand, has more recently also been implicated to cooperate with cell signaling pathways (Huang and Graves, 2003) and seems to inhibit the bacterial sensor NOD2's (nucleotide-binding oligomerization domain 2) antibacterial function in human intestinal epithelial cells (Richmond et al., 2012). These examples may illustrate that our transcriptomic data deserves further attention and more detailed analyses and this type of analysis, using different bioinformatics tools and, finally, functional analyses may contribute to elucidate mechanisms underlying biological processes and diseases.

As already discussed above, assigning correct potency classes to chemicals with weak or intermediate potency seems to be a more general problem. Benigni et al. (2016) presented data showing that even experimental *in vivo* systems, though in general correlating well with human data, perform worse for sensitizers of intermediate potency. They further argue that the protein modification step is the rate-limiting step of the whole sensitization process and that *in vitro* tests targeting other AOP events do not add much information. On the other hand, there is only a weak relationship between the rate constant of MA sensitizers as determined by kinetic profiling with a model peptide and their potency in the LLNA (Natsch et al., 2011). This was described to be linked to the anti-inflammatory effect of MA chemicals by inhibiting NF-κB signaling, which increases with reactivity. However, considering the key events in the skin sensitization AOP, the sensitization process as such can be understood as a continuum and would thus not be characterized only by isolated events. Of course, in reality the process must start with the chemical penetrating the skin. In this context, it should be noted that the concept that a chemical's ability to efficiently penetrate the stratum corneum is crucial for its skin sensitization

capacity and potency has recently been proven wrong (Fitzpatrick et al., 2017). Furthermore, access to lower skin layers may in reality be greatly facilitated by impaired skin barrier function, due to, e.g., wet work and small wounds. Interestingly, many strong sensitizers possess irritant properties that correlate with cytotoxicity, and cytotoxicity in turn seems to contribute to sensitizer potency (Jaworska et al., 2015), which is also reflected in our dataset (data not shown). Cytotoxicity is also connected to the protein reactivity of the chemical: Chemicals that are strongly cysteine-reactive are in general cytotoxic, and thus may interfere with vital enzyme function (Bohme et al., 2009; Jaworska et al., 2015). At the same time, irritant effects can generate danger signals such as extracellular ATP or hyaluronic acid degradation products (Esser et al., 2012; Martin et al., 2011), which may serve to activate DCs and, consequently, to prime naïve T cells. Additionally, other factors such as pre-existing inflammation and co-exposures to other substances, which certainly play a role in allergic sensitization, may be hard to implement in any test system. As *in vitro* assays further have to face the demand of being cost-effective and easy to perform, there are obvious limitations to what can be achieved *in vitro*, but the performances of tests so far are very encouraging (Benigni et al., 2016). Although protein reactivity may be very important, cell-based systems should be capable of recapitulating certain additional events on top of peptide reactivity. Alternative assay performance can most likely be further improved as more mechanistic details of skin sensitization are revealed, which will allow identifying both applicability domains and pitfalls more easily.

In conclusion, we have identified a predictive biomarker signature comprising 52 transcripts for classification of skin sensitizing compounds into CLP groups, as required by current legislation. When challenged with 18 independent test compounds, the assay provided accurate results for 78% of potency predictions. It further identified 11/12 sensitizers correctly, which indicates that it is rather conservative, i.e., avoids false-negative predictions. Since the presented biomarker signature is optimized for potency predictions, and not only for binary hazard classifications, we suggest a possible application for our model within an Integrated Testing Strategy for accurate potency predictions, similar to that suggested by (Jaworska et al., 2015). Importantly, hazard and risk characterization can be provided based on a single experimental campaign and only requires an additional bioinformatical analysis step. In addition, the here discussed results effectively illustrate the flexibility and versatility of the GARD setup. Measuring complete transcriptomes of cells provides the opportunity to perform mode-of-action based studies to identify pathways of sensitization, but also to identify predictive biomarker signatures, which we have previously shown also for binary skin sensitization predictions and respiratory sensitization. In addition, the flexibility of the GARD platform will allow for fine-tuning of prediction models as new groups of chemicals are analyzed. Thus, we here present an initial proof of concept of a potency model targeting the CLP groups, which can be modified and improved as more samples are analyzed and more accurate human reference data emerges.

References

Ade, N., Leon, F., Pallardy, M. et al. (2009). HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: Role of the Keap1/Nrf2 pathway. *Toxicol Sci 107*, 451-460. doi:10.1093/toxsci/kfn243

Albrekt, A. S., Johansson, H., Borje, A. et al. (2014). Skin sensitizers differentially regulate signaling pathways in MUTZ-3 cells in relation to their individual potency. *BMC Pharmacol Toxicol 15*, 5. doi:10.1186/2050-6511-15-5

Anderson, S. E., Siegel, P. D. and Meade, B. J. (2011). The LLNA: A brief review of recent advances and limitations. *J Allergy (Cairo) 2011*, 424203. doi:10.1155/2011/424203

Andreas, N., Caroline, B., Leslie, F. 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. doi:10.1016/j.tiv.2010.12.014

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

Bartz, F., Kern, L., Erz, D. et al. (2009). Identification of cholesterol-regulating genes by targeted RNAi screening. *Cell Metab 10*, 63-75. doi:10.1016/j.cmet.2009.05.009

Basketter, D. A., McFadden, J. F., Gerberick, F. et al. (2009). Nothing is perfect, not even the local lymph node assay: A commentary and the implications for REACH. *Contact Dermatitis 60*, 65-69. doi:10.1111/j.1600-0536.2008.01444.x

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

Behroozy, A. and Keegel, T. G. (2014). Wet-work exposure: A main risk factor for occupational hand dermatitis. *Saf Health Work 5*, 175-180. doi:10.1016/j.shaw.2014.08.001

Benigni, R., Bossa, C. and Tcheremenskaia, O. (2016). A data-based exploration of the adverse outcome pathway for skin sensitization points to the necessary requirements for its prediction with alternative methods. *Regul Toxicol Pharmacol 78*, 45-52. doi:10.1016/j.yrtph.2016.04.003

Bohme, A., Thaens, D., Paschke, A. et al. (2009). Kinetic glutathione chemoassay to quantify thiol reactivity of organic electrophiles – Application to alpha,beta-unsaturated ketones, acrylates, and propiolates. *Chem Res Toxicol 22*, 742-750. doi:10.1021/tx800492x

Breiman, L. (2001). Random forests. *Machine Learning 45*, 5-32. doi:10.1023/a:1010933404324

Brodersen, K. H., Ong, C. S., Stephen, K. E. and Buhmann, J. M. (2010). The Balanced Accuracy and Its Posterior Distribution. International Conference on Pattern Recognition. 3121-3124. doi:10.1109/ICPR.2010.764

Buchmann, A. M., Swaminathan, S. and Thimmapaya, B. (1998). Regulation of cellular genes in a chromosomal context by the retinoblastoma tumor suppressor protein. *Mol Cell Biol 18*, 4565-4576. doi:10.1128/MCB.18.8.4565

Chipinda, I., Hettick, J. M. and Siegel, P. D. (2011). Haptenation: Chemical reactivity and protein binding. *J Allergy (Cairo) 2011*, 839682. doi:10.1155/2011/839682

CLP (2016). European Parliament, Council of the European Union accessed July 13th, 2016. http://echa.europa.eu/sv/regulations/clp/

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

Cottrez, F., Boitel, E., Auriault, C. et al. (2015). Genes specifically modulated in sensitized skins allow the detection of sensitizers in a reconstructed human skin model. Development of the SENS-IS assay. *Toxicol In Vitro 29*, 787-802. doi:10.1016/j.tiv.2015.02.012

Cottrez, F., Boitel, E., Ourlin, J. C. et al. (2016). SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitization potency: Reproducibility and predictivity results from an inter-laboratory study. *Toxicol In Vitro 32*, 248-260. doi:10.1016/j.tiv.2016.01.007

Dearden, J. C., Hewitt, M., Roberts, D. W. et al. (2015). Mechanism-based QSAR modeling of skin sensitization. *Chem Res Toxicol 28*, 1975-1986. doi:10.1021/acs.chemrestox.5b00197

Díaz-Uriarte, R. and Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics 7*, 1-13. doi:10.1186/1471-2105-7-3

Diaz-Uriarte, R. (2007). GeneSrF and varSelRF: A web-based tool and R package for gene selection and classification using random forest. *BMC Bioinformatics 8*, 328. doi:10.1186/1471-2105-8-328

Dimitriadou, E., Hornik, K., Leisch, F. et al. (2011). e1071: Misc functions of the Department of statistics (q1071), TU Wien. R package version 1.6. http://CRAN.R-project.org/package=e1071

Dumont, C., Barroso, J., Matys, I. et al. (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. doi:10.1016/j.tiv.2016.04.008

Esser, P. R., Wolfle, U., Durr, C. et al. (2012). Contact sensitizers induce skin inflammation via ROS production and hyaluronic acid degradation. *PLoS One 7*, e41340. doi:10.1371/journal.pone.0041340

European Parliament (2009). Regulation (EC) No 1223/2009 of the European Parliament and of the Council. No 1223/2009. http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009R1223

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

Fitzpatrick, J. M., Roberts, D. W. and Patlewicz, G. (2017). What determines skin sensitization potency: Myths, maybes and realities. The 500 molecular weight cut-off: An updated analysis. *J Appl Toxicol 37*, 105-116. doi:10.1002/jat.3348

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

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

Gerberick, G. F., Ryan, C. A., Dearman, R. J. and Kimber, I. (2007). Local lymph node assay (LLNA) for detection of sensitization capacity of chemicals. *Methods 41*, 54-60. doi:S1046-2023(06)00135-6

Gonda, A., Gal, I., Szanto, S. et al. (2005). CD44, but not l-selectin, is critically involved in leucocyte migration into the skin in a murine model of allergic dermatitis. *Exp Dermatol 14*, 700-708. doi:10.1111/j.0906-6705.2005.00348.x

Harshman, S. W., Young, N. L., Parthun, M. R. and Freitas, M. A. (2013). H1 histones: Current perspectives and challenges. *Nucleic Acids Res 41*, 9593-9609. doi:10.1093/nar/gkt700

Hartung, T. and Rovida, C. (2009). Chemical regulators have overreached. *Nature 460*, 1080-1081. doi:10.1038/4601080a

Heberle, H., Meirelles, G. V., da Silva, F. R. et al. (2015). InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics 16*, 169. doi:10.1186/s12859-015-0611-3

Higashimoto, T., Chan, N., Lee, Y. K. and Zandi, E. (2008). Regulation of I(kappa)B kinase complex by phosphorylation of (gamma)-binding domain of I(kappa)B kinase (beta) by Polo-like kinase 1. *J Biol Chem 283*, 35354-35367. doi:10.1074/jbc.M806258200

Hoffmann, S. (2015). LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. *ALTEX 32*, 379-383. doi:10.14573/altex.1505051

Hu, J., Wang, G., Liu, X. et al. (2013). Polo-like kinase 1 (PLK1) is involved in toll-like receptor (TLR)-mediated TNF-alpha production in monocytic THP-1 cells. *PLoS One 8*, e78832. doi:10.1371/journal.pone.0078832

Huang, M. and Graves, L. M. (2003). De novo synthesis of pyrimidine nucleotides; emerging interfaces with signal transduction pathways. *Cell Mol Life Sci 60*, 321-336. doi:10.1007/s000180300027

Jaworska, J., Dancik, Y., Kern, P. et al. (2013). Bayesian integrated testing strategy to assess skin sensitization potency: From theory to practice. *J Appl Toxicol 33*, 1353-1364. doi:10.1002/jat.2869

Jaworska, J. S., Natsch, A., Ryan, C. et al. (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*, 2355-2383. doi:10.1007/s00204-015-1634-2

Jockers, J. J. and Novak, N. (2006). Different expression of adhesion molecules and tetraspanins of monocytes of patients with atopic eczema. *Allergy 61*, 1419-1422. doi:10.1111/j.1398-9995.2006.01191.x

Johansson, H., Lindstedt, M., Albrekt, A. S. and Borrebaeck, C. A. (2011). A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal

tests. *BMC Genomics 12*, 399. doi:10.1186/1471-2164-12-399

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

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

Johnson, P. and Ruffell, B. (2009). CD44 and its role in inflammation and inflammatory diseases. *Inflamm Allergy Drug Targets 8*, 208-220. doi:10.2174/187152809788680994

Johnson, W. E., Li, C. and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics 8*, 118-127. doi:10.1093/biostatistics/kxj037

Kohle, C. and Bock, K. W. (2007). Coordinate regulation of Phase I and II xenobiotic metabolisms by the Ah receptor and Nrf2. *Biochem Pharmacol 73*, 1853-1862. doi:10.1016/j.bcp.2007.01.009

KPA (2016). Key Pathway Advisor by Clarivate Analytics (Formerly the IP & Science business of Thomson Reuters). http://ipscience.thomsonreuters.com/product/metacore/.

Lane, A. N. and Fan, T. W. (2015). Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Res 43*, 2466-2485. doi:10.1093/nar/gkv047

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

Lee-Sayer, S. S., Dong, Y., Arif, A. A. et al. (2015). The where, when, how, and why of hyaluronan binding by immune cells. *Front Immunol 6*, 150. doi:10.3389/fimmu.2015.00150

Leek, J. T., Johnson, W. E., Parker, H. S. et al. (2014). sva: Surrogate Variable Analysis. R package version 3.10.0.

Li, S. X., Tong, Y. P., Xie, X. C. et al. (2007). Octameric structure of the human bifunctional enzyme PAICS in purine biosynthesis. *J Mol Biol 366*, 1603-1614. doi:10.1016/j.jmb.2006.12.027

Liaw, A. and Wiener, M. (2002). Classification and regression by random forest. *R News 2,* 18-22.

Lu, X., Li, Y., Wang, W. et al. (2014). 3 beta-hydroxysteroid-Delta 24 reductase (DHCR24) protects neuronal cells from apoptotic cell death induced by endoplasmic reticulum (ER) stress. *PLoS One 9*, e86753. doi:10.1371/journal.pone.0086753

Luechtefeld, T., Maertens, A., McKim, J. M. et al. (2015). Probabilistic hazard assessment for skin sensitization potency by dose-response modeling using feature elimination instead of quantitative structure-activity relationships. *J Appl Toxicol 35*, 1361-1371. doi:10.1002/jat.3172

Magnusson, B. and Kligman, A. M. (1969). The identification of contact allergens by animal assay. The guinea pig maximization test. *J Invest Dermatol 52*, 268-276. doi:10.1038/jid.1969.42

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

Migdal, C., Botton, J., El Ali, Z. et al. (2013). Reactivity of chemical sensitizers toward amino acids in cellulo plays a role in the activation of the Nrf2-ARE pathway in human monocyte dendritic cells and the THP-1 cell line. *Toxicol Sci 133*, 259-274. doi:10.1093/toxsci/kft075

Natsch, A. and Emter, R. (2008). Skin sensitizers induce antioxidant response element dependent genes: Application to the in vitro testing of the sensitization potential of chemicals. *Toxicol Sci 102*, 110-119. doi:10.1093/toxsci/kfm259

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

Natsch, A., Gfeller, H., Kuhn, F. et al. (2010). Chemical basis for the extreme skin sensitization potency of (E)-4-(ethoxymethylene)-2-phenyloxazol-5(4H)-one. *Chem Res Toxicol 23*, 1913-1920. doi:10.1021/tx1002707

Natsch, A., Haupt, T. and Laue, H. (2011). Relating skin sensitizing potency to chemical reactivity: Reactive Michael acceptors inhibit NF-kappaB signaling and are less sensitizing than S(N)Ar- and S(N)2-reactive chemicals. *Chem Res Toxicol 24*, 2018-2027. doi:10.1021/tx2003678

Natsch, A., Emter, R., Gfeller, H. et al. (2015). Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: Global versus domain-based assessment. *Toxicol Sci 143*, 319-332. doi:10.1093/toxsci/kfu229

OECD (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. 1-59.

Patlewicz, G., Jeliazkova, N., Safford, R. J. et al. (2008). An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ Res 19*, 495-524. doi:10.1080/10629360802083871

Peiser, M., Tralau, T., Heidler, J. 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. doi:10.1007/s00018-011-0846-8

Peng, W. M., Yu, C. F., Kolanus, W. et al. (2011). Tetraspanins CD9 and CD81 are molecular partners of trimeric FcεRI on human antigen-presenting cells. *Allergy 66*, 605-611. doi:10.1111/j.1398-9995.2010.02524.x

Piccolo, S. R., Sun, Y., Campbell, J. D. et al. (2012). A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics 100*, 337-344. doi:10.1016/j.ygeno.2012.08.003

Piccolo, S. R., Withers, M. R., Francis, O. E. et al. (2013). Multiplatform single-sample estimates of transcriptional activation. *Proc Natl Acad Sci U S A 110*, 17778-17783. doi:10.1073/pnas.1305823110

Piroird, C., Ovigne, J. M., Rousset, F. et al. (2015). The myeloid U937 skin sensitization test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome

pathway for skin sensitization. *Toxicol In Vitro 29*, 901-916. doi:10.1016/j.tiv.2015.03.009

Qiu, G., Sun, W., Zou, Y. et al. (2015). RNA interference against TMEM97 inhibits cell proliferation, migration, and invasion in glioma cells. *Tumour Biol 36*, 8231-8238. doi:10.1007/s13277-015-3552-6

Richmond, A. L., Kabi, A., Homer, C. R. et al. (2012). The nucleotide synthesis enzyme CAD inhibits NOD2 antibacterial function in human intestinal epithelial cells. *Gastroenterology 142*, 1483-1492.e6. doi:10.1053/j.gastro.2012.02.040

Roberts, D. W., Aptula, A. O. and Patlewicz, G. (2007). Electrophilic chemistry related to skin sensitization. Reaction mechanistic applicability domain classification for a published data set of 106 chemicals tested in the mouse local lymph node assay. *Chem Res Toxicol 20*, 44-60. doi:10.1021/tx060121y

Roggen, E. L. and Blaauboer, B. J. (2013). Sens-it-iv: A European Union project to develop novel tools for the identification of skin and respiratory sensitizers. *Toxicol In Vitro 27*, 1121. doi:10.1016/j.tiv.2013.01.009

Schulz, V. J., van Roest, M., Bol-Schoenmakers, M. et al. (2013). Aryl hydrocarbon receptor activation affects the dendritic cell phenotype and function during allergic sensitization. *Immunobiology 218*, 1055-1062. doi:10.1016/j.imbio.2013.01.004

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

Singh, R., Mortazavi, A., Telu, K. H. et al. (2013). Increasing the complexity of chromatin: Functionally distinct roles for replication-dependent histone H2A isoforms in cell proliferation and carcinogenesis. *Nucleic Acids Res 41*, 9284-9295. doi:10.1093/nar/gkt736

Teunis, M. A., Spiekstra, S. W., Smits, M. et al. (2014). International ring trial of the epidermal equivalent sensitizer potency assay: Reproducibility and predictive-capacity. *ALTEX 31*, 251-268. doi:10.14573/altex.1308021

Tsujita-Inoue, K., Hirota, M., Ashikaga, T. et al. (2014). Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays. *Toxicol In Vitro 28*, 626-639. doi:10.1016/j.tiv.2014.01.003

Urbisch, D., Mehling, A., Guth, K. et al. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul Toxicol Pharmacol 71*, 337-351. doi:10.1016/j.yrtph.2014.12.008

Waterham, H. R., Koster, J., Romeijn, G. J. et al. (2001). Mutations in the 3beta-hydroxysterol Delta24-reductase gene cause desmosterolosis, an autosomal recessive disorder of cholesterol biosynthesis. *Am J Hum Genet 69*, 685-694. doi:10.1086/323473

Xu, Z., Cetin, B., Anger, M. et al. (2009). Structure and function of the PP2A-shugoshin interaction. *Mol Cell 35*, 426-441. doi:10.1016/j.molcel.2009.06.031

Conflict of interest

The authors declare no competing financial interest. A patent application has been submitted related to the content of this article.

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 Ann-Charlott Olsson for microarray sample preparation and Cosmetics Europe – The Personal Care Association for providing 27 of the chemicals used in this study.

Correspondence to

Malin Lindstedt, PhD
Lund University
Medicon Village (406)
22381 Lund, Sweden
Phone: +46 46 2229256
Fax: +46 46 2224200
e-mail: malin.lindstedt@immun.lth.se