



An alternative biomarker-based approach for the prediction of proteins known to sensitize the respiratory tract



Kathrin S. Zeller^a, Henrik Johansson^b, Trine Østergaard Lund^c, Nanna N. Kristensen^c,
Erwin L. Roggen^c, Malin Lindstedt^{a,*}

^a Department of Immunotechnology, Lund, Sweden

^b SenzaGen AB, Lund, Sweden

^c Novozymes A/S, Bagsvaerd, Denmark

ARTICLE INFO

Keywords:

Respiratory sensitization
Protein allergens
Cell-based assay
Dendritic cells
Biomarker

ABSTRACT

Many natural and industrial proteins are known to have properties that can result in type I hypersensitivity, however, to date, no validated test system exists that can predict the sensitizing potential of these allergens. Thus, the objective of this study was to develop a protocol based on the myeloid cell-based Genomic Allergen Rapid Detection (GARD) assay that can be used to assess and predict the capacity of protein allergens known to induce sensitization in the respiratory tract. Cellular responses induced by eight selected proteins were assessed using transcriptional profiling, flow cytometry and multiplex cytokine analysis. 391 potential biomarkers were identified as a predictive signature and a series of cross-validations supported the validity of the model. These results together with biological pathway analysis of the transcriptomic data indicate that the investigated cell system is able to capture relevant events linked to type I hypersensitization.

1. Introduction

Allergy is a chronic disease with increasing prevalence and it is of outmost importance for the industry and authorities to identify potential allergens as early as possible to limit the exposure of workers and the general populations. Several hundreds of chemicals are known to be able to cause allergic contact dermatitis (DeGroot, 2008; Luechtefeld et al., 2016), a type IV delayed hypersensitivity reaction, whereas less chemicals are known to sensitize the respiratory tract and to induce type I allergic responses (Boverhof et al., 2008). Most substances causing respiratory allergy are proteins of environmental origin e.g. allergens from house dust mite feces, pollen, or fungi, while others are present in an occupational setting such as enzymes used in flavor, fragrance, detergents and pharmaceutical production (Thomas et al., 2002; Baur, 2005). The risk of developing adverse reactions following occupational exposure exists; thus, a strict focus on occupational safety is mandatory. Sensitization has been observed for workers exposed to certain industrial enzymes such as α -amylase, proteases, pancreatin, and trypsin (Baur et al., 2013; Budnik et al., 2017). New enzymes are continuously developed for existing as well as for new applications, such as genetically modified enzymes used in food processing and flavor production and may also lead to occupational health risks (Baur, 2005; Budnik et al., 2016).

To date, no validated assay is available for predicting the sensitizing potential of novel proteins, rendering a weight-of-evidence approach to be the most acceptable means of allergy safety assessment. There is, however, a growing consensus that the allergenic potential of compounds, including proteins, should be evaluated with regard to their biochemical characteristics and the protein's potential to induce a specific immune response (European COST Project impARAS (<http://imparas.eu/>, n.d.)). A combination of physical traits of proteins, the molecular interaction between human cells and proteins, as well as their impact on cell-cell interactions play a role in understanding and eventually predicting protein allergenicity (McClain et al., 2014; North et al., 2016).

The Adverse Outcome Pathway (AOP) concept is a framework for collecting and organizing information relevant to an adverse outcome at different levels of biological organization (OECD et al., 2014). These AOPs are based on available information on substance-response and response-response relationships and allow the development of relevant predictive animal-free test methods and approaches, as well as the contextualization of the results across a diverse range of biological mechanisms and toxicity endpoints. The mechanisms driving respiratory sensitization are not fully disclosed yet, but emerging data suggest that events driving respiratory (chemical- and protein-induced) and skin sensitization (chemical-induced) can be structured in the same

* Corresponding author at: Lund University, Medicon Village (406), 22381 Lund, Sweden.
E-mail address: malin.lindstedt@immun.lth.se (M. Lindstedt).

adverse outcome pathway (AOP). In contrast to skin sensitization, properly evaluated test methods addressing the key events of respiratory sensitization induction are not yet available (Roggen, 2014).

The Genomic Allergen Rapid Detection (GARD) assay was developed to provide information about the capacity of chemicals to induce skin sensitization (accuracy: 89% (Johansson et al., 2014)). This *in vitro* assay utilizes a myeloid cell line resembling dendritic cells (DCs) as a model system. DCs are antigen-presenting cells and central for the induction and regulation of adaptive immune responses (Schoorhuis et al., 2006). This assay was recognized by both the European Reference Laboratory – European Center for Validation of Alternative Methods (EURL-ECVAM) and the OECD as a valuable method for addressing key event 3 (Dendritic cell activation and maturation) of the AOP for skin sensitization (OECD, 2012). Since DCs bridge the innate and adaptive immunity, a requirement of both skin and respiratory sensitization to occur, it can be hypothesized that this cell system may also be suited for the identification of chemicals causing sensitization in the respiratory tract. Forreryd et al. (Forreryd et al., 2015) successfully demonstrated that a modified protocol of the assay is able to predict respiratory chemical sensitizers with an accuracy of 84% based on a biomarker signature consisting of 389 transcripts.

In this study, this protocol is assessed for its capacity to provide a mechanistic understanding of the protein sensitization at DC level. The objective was to identify a potential gene profile for the classifications of proteinaceous allergens. For this purpose, the effect of eight respiratory protein allergens on the gene expression in the myeloid cell line was investigated using Affymetrix RNA expression array technology. A panel of potential biomarkers distinguishing the allergens from experimental control samples was identified using data-driven approaches including machine learning, and cross-validation exercises indicated that the identified transcripts may indeed be useful for classification. These biomarkers were further investigated with regard to associated biological pathways.

2. Materials and methods

2.1. Respiratory allergens

All allergens contain low levels of endotoxin as summarized in Table 1. LoTox™ recombinant Der p 7 (#LTR-DP7-1) and LoTox™ natural Der p 1 (#LTN-DP1-1) were obtained from Indoor biotechnologies, Charlottesville, USA. The other allergens were provided by Novozymes A/S, Bagsvaerd, Denmark, and tested for endotoxin content by Sahlgrenska Universitetssjukhuset, Bakteriologiska laboratoriet, Göteborg.

2.2. Cell culture and stimulations

The maintenance of the cell line was performed as described in (Johansson et al., 2011). In short, the MUTZ-3 derivative cells were

cultured in MEM alpha modification (Nordic Biolabs/GE Healthcare Bio-Sciences, Täby, Sweden) supplemented with 20% fetal bovine serum (Thermo Fisher Scientific/Life Technologies, Stockholm, Sweden) and 40 ng/mL recombinant human GM-CSF (Miltenyi Biotec Norden AB, Lund, Sweden). A dose finding experiment was performed based on earlier protocols optimized for this cell line to identify the highest enzyme concentration resulting in a relative cell viability of $\geq 90\%$ after 24 h of incubation. Longer exposures (48 h) to certain enzymes resulted in substantially decreased cell viability (data not shown). A phenotypic control analysis of cells prior to each experiment was carried out by flow cytometry (see below) in order to assure an immature state. In short, three batches of cells were exposed to allergens and control substances dissolved in complete cell culture medium for 24 h in at least three independent experiments (Table 1, six independent experiments for “unstim”). Proteins present in serum served as vehicle and non-allergen protein controls. Phenotypic controls and viability of the cells were assessed after the stimulation period by analyzing cell surface expression using flow cytometry as described below. Cells aimed for RNA extraction were lysed in TRIzol® (Life Technologies/Thermo Fisher Scientific, Waltham, USA) and stored until further use in minus 20 °C.

2.3. Flow cytometry

As a quality control both before and during the experiments, the following monoclonal antibodies were used during phenotypic analysis as described previously (Johansson et al., 2013): CD1a (DakoCytomation, Glostrup, Denmark), CD34, CD86, HLA-DR (BD Biosciences, Franklin Lakes, NJ), all FITC-conjugated; CD14 (DakoCytomation), CD54, CD80 (BD Biosciences), all PE-conjugated. FITC- and PE-conjugated mouse IgG1 (BD Biosciences) were used as isotype controls and Propidium iodide was included as a marker for non-viable cells (BD Biosciences). After 24 h incubation with the allergens and vehicle control samples, viability, CD14, CD1a, HLA-DR and CD86 expression were assessed. 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). An endotoxin stimulation was included as internal control for cell quality.

2.4. RNA extraction, cDNA and array hybridization

RNA isolation from cells lysed in TRIzol® was performed according to manufacturer's instructions. Labeled sense DNA was synthesized according to Affymetrix' protocols using the recommended kits and controls. The cDNA was hybridized to Human Gene 1.0 ST arrays and further processed and scanned as recommended by the supplier (Affymetrix, Cleveland, USA).

2.5. Data acquisition and analysis

The dataset, comprising 33 samples in total, was quality-controlled and then normalized using Single Channel Array Normalization (Piccolo et al., 2012; Piccolo et al., 2013). Statistical analysis was performed primarily by Analysis of Variation (ANOVA). ANOVAs and PCA visualization of results were performed with Qlucore 3.2 (Qlucore AB, Lund, Sweden). Significance was evaluated with the multiple-hypothesis corrected p-value (q-value, in this article referred to as false discovery rate (FDR) (Benjamini and Hochberg, 1995)). An $FDR \leq 0.05$ was considered as statistically significant. Decision values were calculated in a support vector machine (SVM) model (Noble, 2006), constructed in R (R Development Core Team, 2008) and using the package e1071 (R package e1071). The same R package was used to program the backward elimination algorithm (Carlsson et al., 2011). The backward elimination algorithm was applied to a dataset consisting of nine vehicle control and 24 allergen samples and the 1052 most significant

Table 1
Identity and concentrations of allergens and vehicle controls.

Substance	Test compound concentration	Endotoxin in well conc. [EU/mL]
Amylase 1	25 µg/mL	0.04
Amylase 2	25 µg/mL	0.059
Protease 1	25 µg/mL	< 0.0012
Protease 2	25 µg/mL	< 0.0006
Glycohydrolase	25 µg/mL	0.033
Lipase	25 µg/mL	0.022
Der p 1	25 µg/mL	< 0.75
Der p 7	25 µg/mL	< 0.25
DMSO	0.1%	NA
Unstimulated		NA

*Unstimulated cells are defined as cells grown in control medium, i.e. not exposed to any potentially stimulating substance.

genes based on a two group comparison between allergens and controls ($p = 0.007$, $FDR = 0.19212$). Cross-validations were performed in a similar manner based on an input where one treatment at a time had been removed. All additional statistical computing and handling of data was performed in Excel (Microsoft Office 2013) and R (www.R-project.org).

2.6. Bio-Plex 200 cytokine analysis

Cell supernatants were analyzed using a BioPlex 200 system (Bio-Rad, Hercules, USA) with the Novex® Human Cytokine Magnetic 30-Plex Panel kit (# LHC6003M, Invitrogen/Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's recommendations. Biological triplicates, based on three batches of cells, were analyzed in technical duplicates (maximum CV accepted was 20%) and concentrations in range were accepted according to the Bio-Plex 200 Manager 6.1 software (Bio-Rad, Hercules, USA) unless stated otherwise. Data are presented as mean values; error bars represent standard deviations. Due to a limited number of samples (biological triplicates, technical duplicates) and non-normally distributed data, p-values as a measure of statistical significance are not appropriate to provide. An endotoxin stimulation was included as internal assay control.

2.7. Key pathway analysis

The Key Pathway Advisor (KPA, 2016) tool versions 16.4 and 16.6, were used to investigate the identified 391 genes comprising the protein prediction signature in a biological context. KPA associates e.g. differentially expressed genes derived from RNA microarray data with both upstream and downstream processes in order to allow biological interpretation. The investigated data set consisted of 33 samples, comprising allergen and control samples in at least three replicates. Both p values and fold changes were used as input based on a two group comparison between allergens and vehicle controls (Key Hubs Calculation Algorithm: causal reasoning analysis; Key Hubs p-value threshold = 0.01; Key Processes p-value threshold = 0.05; unrecognized IDs: 105). An analysis using the same parameters was performed with 272 genes as input, which resulted as the most significant genes from a p-value filtering of proteases versus unstimulated samples ($p = 0.001335$, $FDR = 0.05$) based on a variance-prefiltered data set containing 10,054 variables. 254 key hubs were predicted and nine IDs were not recognized.

3. Results

3.1. Statistical analysis of expression array data

The data set consisted of 33 samples in total, nine experimental controls and 24 allergen samples (Table 1). When comparing these two groups (i.e. controls, “unstimulated” and “DMSO”, versus allergens), no significantly regulated genes could be detected based on a false discovery rate (FDR) of 0.05. With a chosen cut-off of $p = 0.0001$ ($FDR = 0.0847$), 31 variables were left. This indicates that only smaller differences exist between the group of protein allergens and experimental controls containing serum proteins. When comparing all treatments to each other (multigroup ANOVA, $FDR = 0.01$), 1037 significantly regulated genes were identified. All samples cluster rather tightly except the proteases as presented in a Principal Component Analysis (PCA) plot in Fig. 1, indicating that the RNA expression patterns induced by each protease differ somewhat from all the other samples.

3.2. Identification of a protein allergen prediction signature

In order to develop a classification approach, a so-called wrapper method was used to identify the most relevant genes for distinguishing

allergens from non-allergens after initial p-value filtering. The used backward elimination algorithm (Carlsson et al., 2011) based on support vector machine (SVM) predictions (Cortes and Vapnik, 1995) results in an optimized list of genes. Each gene or feature in this list has been evaluated not only for its importance for the classification itself, but also how it performs together with the other features. Using an input of the most significant 1052 genes, based on a two-group comparison (allergen versus non-allergen, $p = 0.007$, $FDR = 0.19$), the algorithms identified a signature consisting of 391 genes (Supplementary Table 1). When the dataset is visualized by PCA using these potential biomarkers as an input, a clear separation between allergens and experimental controls can be seen (Fig. 2).

In order to estimate the accuracy of the model and to ensure that the presented results were not achieved by chance, a series of leave-one-out cross-validations was performed. For each cross-validation exercise, all replicates of one stimulation were removed and the backward elimination procedure was repeated as described above. The obtained gene signature was then used to predict the left out samples using an SVM model based on a training set consisting of the remaining samples in the dataset. All stimulations were removed once except “unstimulated”; without the “unstimulated” samples, the dataset would be too unbalanced to be used for training of a model. The resulting SVM predictions are visualized as boxplots in Fig. 3.

All allergens were correctly predicted as such by majority voting, i.e. the alternative with the majority of predictions, corresponding to more than half of the replicate samples, was selected. One replicate each for lipase and Der p 7 was predicted as non-allergen (negative decision value). SVM predictions were also performed after a similar exercise based on a dataset where DMSO and Der p 1 samples were left out and subsequently used as a test set. Also here, DMSO was correctly classified as non-allergen and Der p 1 as an allergen (data not shown).

3.3. Biological pathway analysis reveals association of gene lists with immunological pathway regulation

The association of the 391 identified potential biomarkers with biological pathways were investigated using the Key Pathway Advisor tool, and the gene list together with associated p-values and fold changes based on a two-group comparison (allergen yes/no) were used as input. Of in total 27 significant pathways identified (Fig. 4), roughly one third was allocated to immune system-related pathways, e.g. IL-3, IL-5, IL-6, IL-15, IL-17 and IL-18 signaling pathways. Examples of suggested key hubs, all with increased activity, were MHC class I, NF κ B variants p50 and p50/p65, IL-10, IL-6R, IL-10R, Fc γ RII α , glutaredoxin, and integrins $\alpha\beta_4$ and α_{2B} .

When investigating biological pathways based on a comparison between proteases and unstimulated samples and the resulting 272 most significant genes, 59 significant pathways were identified and over 40% are related to immune responses as judged by their name (Fig. 5), e.g. IL-2, IL-10, IL-33 and CD40 signaling. Further, “CCR4-induced chemotaxis of immune cells”, “CCR4-dependent immune cell chemotaxis in asthma and atopic dermatitis”, and NF κ B-related pathways are associated to inflammatory responses. NF κ B also appeared as a predicted key hub, along with other molecules such as MyD88, an adaptor protein used by almost all TLRs to activate NF κ B. Toll-like receptor 4 and high-mobility group box-1 (HMGB1), a transcriptional regulator involved in inflammation and cell differentiation, were two more examples of predicted key hubs.

3.4. Protein allergens affect both surface marker expression and secreted cytokines

As a complement to the transcriptional analysis, effects on the protein level were investigated by FACS, focusing on a set of differentiation and activation markers as described in Fig. 6. All cells were positive for HLA-DR in the unstimulated control as well as after

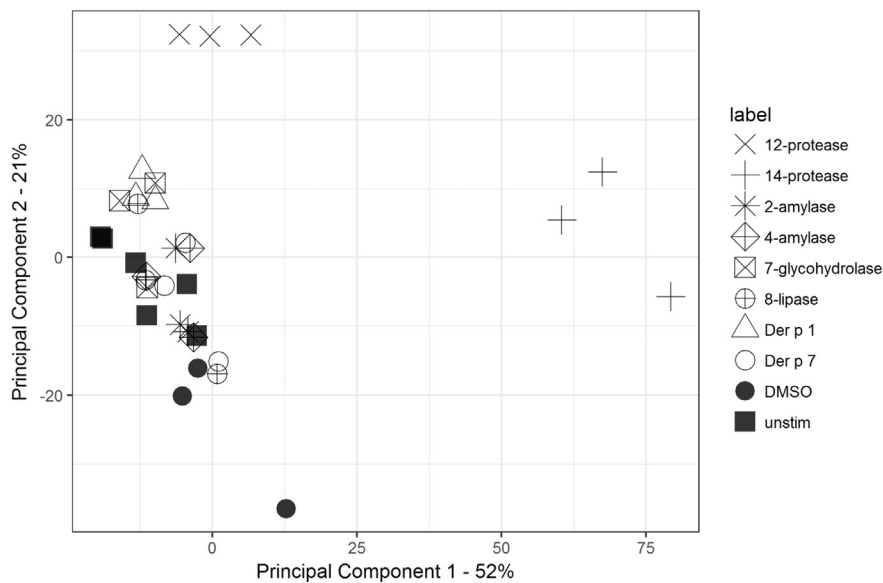


Fig. 1. Principal component analysis plot visualizing differences in transcript expression levels based on at least biological triplicate samples treated with the indicated substances (multi-group ANOVA comparing treatments; FDR = 0.01, 1037 genes. Unstim = control cells “treated” only with (vehicle) cell culture medium). Please note that one unstim sample is not visible in this plot due to an overlay.

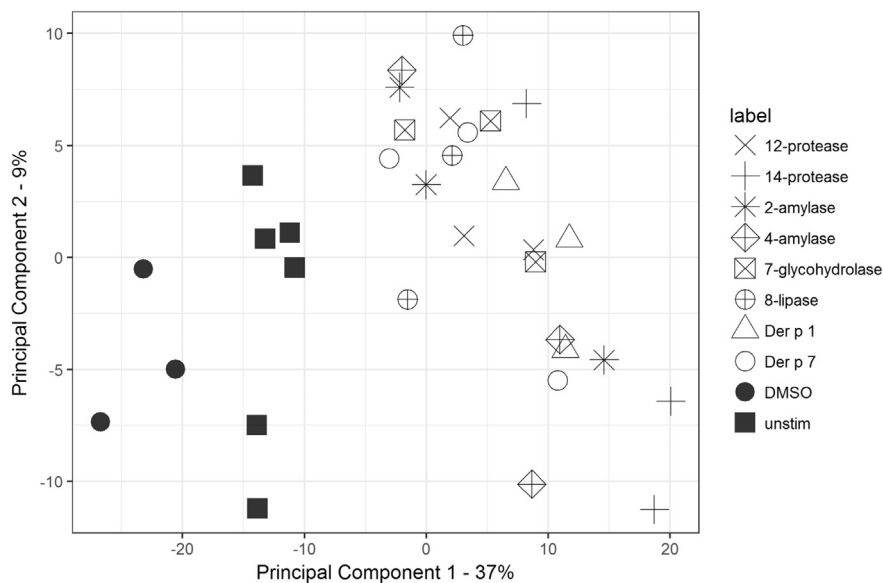


Fig. 2. Principal component analysis plot visualizing differences in transcript expression levels of all 33 samples (at least biological triplicates treated with the indicated substances) based on an input of the identified 391 potential biomarkers. Unstim = control cells “treated” only with (vehicle) cell culture medium.

exposure to all substances (data not shown). There was no protein allergen that reduced the presence of any investigated marker; however, protease 1 seemed to increase expression of e.g. CD86, a known activation marker. Notably, the relative viability after all treatments was $\geq 90\%$ (Suppl. Fig. S1).

When screening 30 soluble cytokines using Bio-Plex® technology, Der p 1 and amylase 1 treatment increased the levels of MIP-1 α and Rantes in supernatants (Fig. 7). The proteases reduced the levels of several cytokines; likely at least for some by degradation as supported by the results for GM-CSF. GM-CSF is added to the cell medium at 40 ng/mL, but after incubation with the proteases, the values were lowered to a few pg/mL (protease 2) and to approximately 500 pg/mL (protease 1), respectively (Suppl. Fig. S2), whereas the values of all the other allergens and controls were as expected out of range at the upper limit.

However, the proteases increased the levels of e.g. IL-8 (Fig. 8), and protease 1 treatment induced higher levels of Rantes (Fig. 7A), IL-12, MCP-1 and IL-1RA levels (Fig. 8). The data on cytokine protein expression were in part correlated with the changes on RNA level (Suppl. Fig. S3).

4. Discussion

In this article, we describe the development of an *in vitro* method for the prediction of protein allergens using eight known respiratory allergens as model substances. Using data-driven approaches based on machine learning, we identified a list of potential biomarkers, which we further investigated with regard to associated biological pathways.

Comparing the RNA expression pattern of cells exposed to experimental controls and those exposed to the respiratory protein allergens, no significantly regulated transcripts based on an FDR of 0.05 could be observed. This could be due to several factors, e.g. relatively small differences between the groups, but also high variation in the transcriptional changes induced by the structurally and functionally different enzymes. Unexpectedly, even when comparing the intended positive controls, Der p 1 and Der p 7, separately to the vehicle samples, no significantly regulated genes could be identified. On the protein level, however, Der p 1 seemed to increase Rantes and MIP-1 α cytokine levels (Fig. 7), which would be consistent with a pro-inflammatory response induced by Der p 1 (Konig et al., 2015; King et al., 1998). As endotoxins even at low concentrations may activate cells it is important to consider their effects as they may obliterate allergen signals. Der p 1

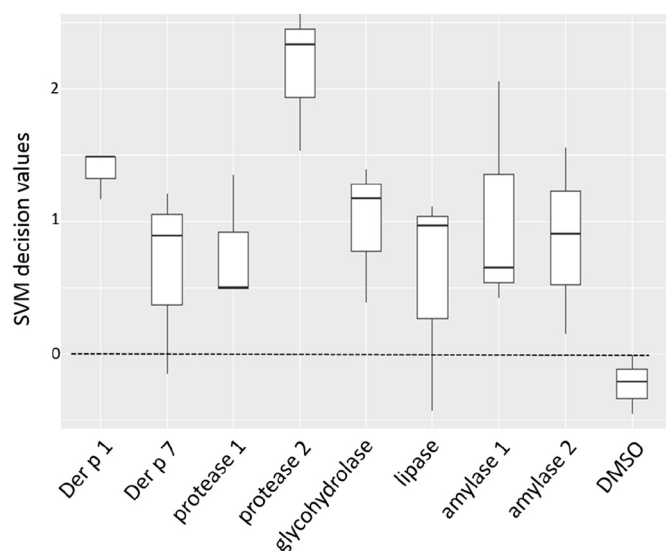


Fig. 3. A series of leave-one-out cross-validations was performed. For each cross-validation exercise, all replicates of the indicated treatment were removed and the backward elimination procedure was performed as described in Materials and Methods. Resulting predictions for the respective treatment are presented as boxplots depicting Support vector machine (SVM) decision values for each replicate. Negative DVs correspond to non-allergens, positive DVs to allergens.

and Der p 7 have therefore been obtained as “low endotoxin” preparations; nevertheless they contain the highest amounts of endotoxin of all samples considering final concentrations (Table 1, 0.75 EU/mL and 0.25 EU/mL, respectively). As comparison, standard endotoxin levels in fetal bovine serum usually account to ≤ 1 ng/mL or ≤ 10 EU/mL, respectively, according to the industrial standard (http://www.serumsourceintl.com/questions_answers.html, 2015). However, endotoxins are likely also an intrinsic part of allergic reactions in “real life”, as endotoxins are more or less ubiquitously present. Furthermore, Der p 1 preparations, including the product used here, commonly do not exhibit cysteine protease activity due to the production process. For Der p 1, cysteine protease activity has been described to enhance production of Der p 1-specific IgE, and allergenicity in a mouse inhalation model (Gough et al., 2003; Gough et al., 1999). On the other hand, data presented by Porter et al. (Porter et al., 2009) indicate that Der p 1 in house dust samples does not have measurable protease activity. In the here presented study, Der p 1's cysteine protease activity was not restored by reducing treatment, as we could not detect any difference in activation of monocyte-derived dendritic cells when comparing cells stimulated with Der p 1 treated with DTT in order to activate its cysteine protease activity with non-treated Der p 1 (part of an unpublished study, data not shown). Nevertheless, the identified biomarker signature consisting of 391 transcripts was able to separate the dataset into two clearly distinct groups, i.e. experimental controls and allergens (Fig. 2). Though separate statistical analyses could not identify significantly regulated transcripts between these two groups, this together with the cross-validation results (Fig. 3) indicates that several

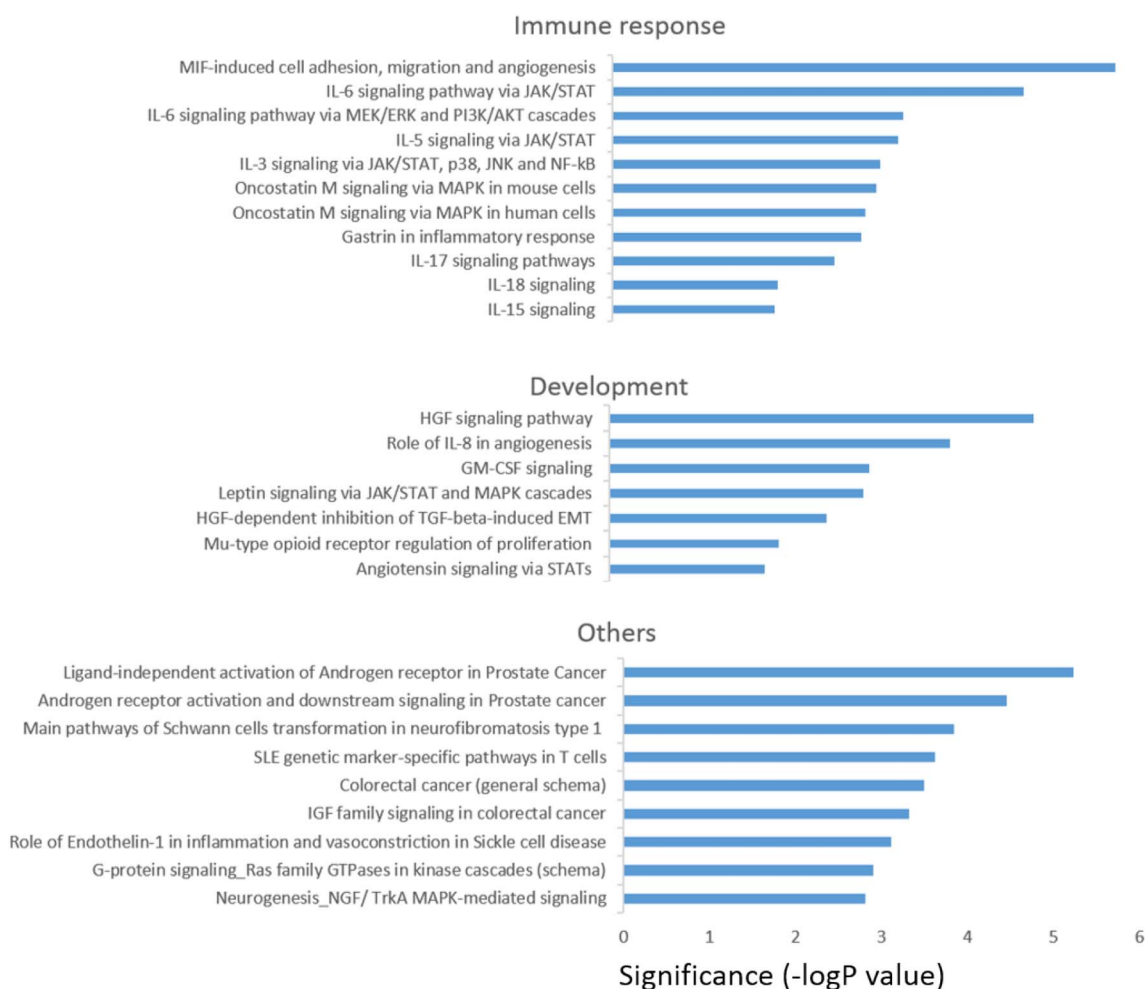


Fig. 4. Key Pathway Advisor analysis results in 27 significantly regulated pathways using the 391 transcripts as input, which were identified by a two group comparison between six protein allergens and controls followed by the application of a backward elimination algorithm.

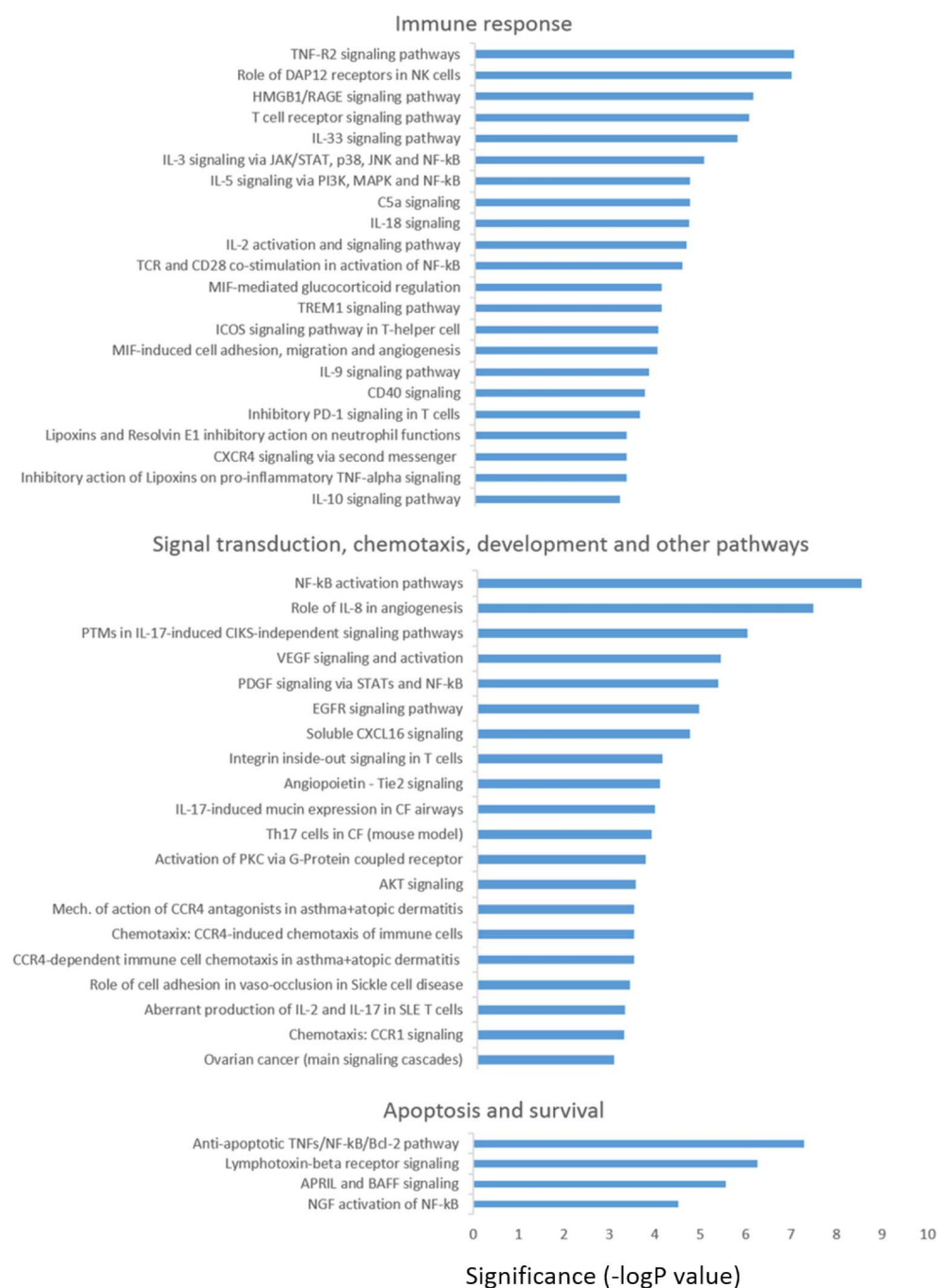


Fig. 5. Key Pathway Advisor analysis based on an input of the most significant 272 transcripts resulting from a two-group comparison of unstimulated control samples and protease-treated samples. The most significant pathways are presented (Union p-values ≤ 0.001 , corresponding to a $-\log$ p-value of 3).

transcripts together in a signature may still be useful for the prediction of the protein allergens capable of sensitizing the respiratory tract.

Biological pathway analysis using the KPA tool and the potential biomarker signature as input, revealed an overrepresentation of numerous pathways linked to immune responses. Further evidence supporting the relevance of the identified pathways for respiratory sensitization is provided by the current understanding of the NF-κB signaling pathway, which was a suggested upstream regulator for the identified transcriptional changes. NF-κB activation is causally related to increased release of various signal molecules such as IL-33, IL-25 (Corsini et al., 2013; Takai et al., 2014), endogenous danger factors (e.g. HMGB-1, uric acid and ATP by epithelial cells and DCs); further to DC activation and migration (Roggen et al., 2014), and the induction of ovalbumin (Ather et al., 2011) and Der p 2 sensitization (Chiou and Lin, 2009). The release of ATP and uric acid drives the activation of the

NLRP3 inflammasome complex resulting in cleavage of pro-IL-1β to mature IL-1β through caspase 1 (Yazdi et al., 2007). Uric acid may play an important role in Th2 skewing (Kool et al., 2011).

Although human and animal data indicate that the here investigated protein allergens do not differ significantly in terms of allergenicity (Basketter et al., 2010), in our *in vitro* system, protease 1 and 2 induced a particular RNA expression pattern and also influenced the protein levels of certain cell surface markers and cytokines. This may not necessarily be associated with their allergenicity, as it could also be related to cytotoxic and adjuvant properties as described for several cysteine and serine proteases and certain proteases used in detergents (Sarlo et al., 1997; Scheurer et al., 2015; Schweigert et al., 2000; Chapman et al., 2007). When the cells were exposed to the proteases for 48 h, relative cell viability decreased (data not shown), however, the relative viability never dropped below 90% after 24 h (Fig. S2).

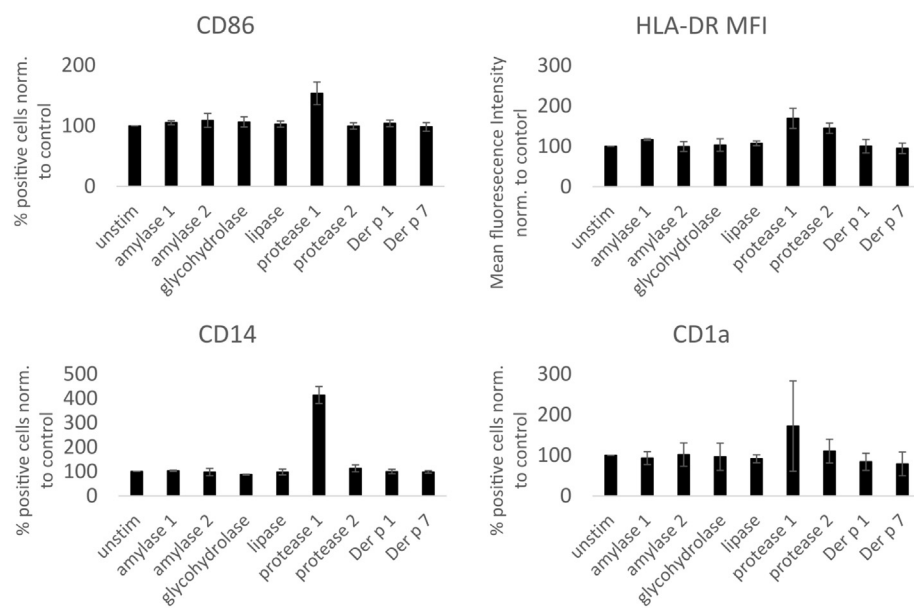


Fig. 6. Changes in cell surface expression levels of CD86, CD14, CD1a, and HLA-DR in MUTZ-3 cells stimulated with the indicated substances for 24 h as determined by flow cytometry ($n = 3$, error bars show standard deviations). Data are normalized to respective control (unstimulated) sample (control = 100%). MFI = mean fluorescence intensity.

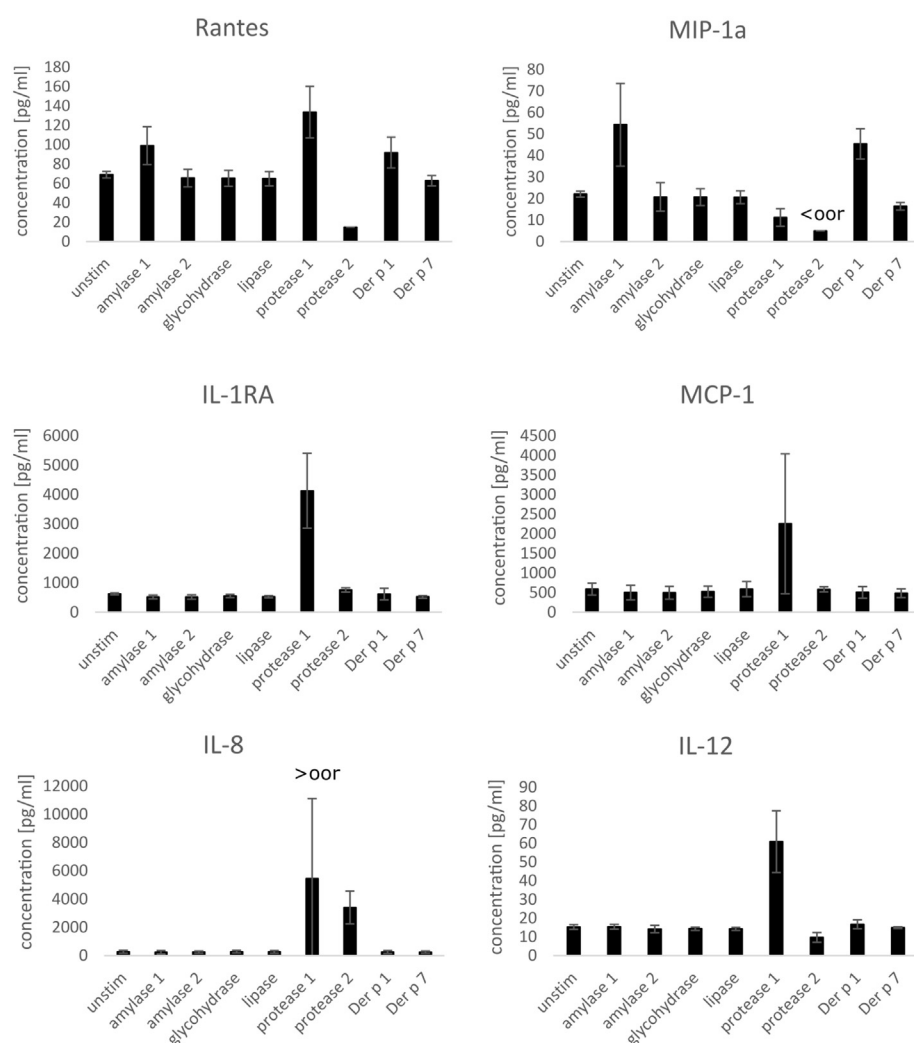


Fig. 7. Changes in Rantes and MIP-1 α protein levels in MUTZ-3 supernatants after stimulation with the indicated substances for 24 h, as determined by Bio-Plex® technology. ($n = 3$, error bars show standard deviations.)

Fig. 8. Changes in protein levels of the indicated cytokines in MUTZ-3 supernatants after stimulation with the indicated substances for 24 h, as determined by Bio-Plex®. ($n = 3$, error bars show standard deviations.)

Although it is expected that proteases are able to degrade certain proteins such as cytokines as shown for GM-CSF (Fig. S3) and receptors (as reviewed in (Straube et al., 2005)), based on the here presented investigations, it is not possible to make conclusions about which

mechanisms are involved. While the multiplex cytokine assay has been of value in this discovery study, complementing the picture drawn by transcriptomic measurements, it is clear that the proteases may degrade some of the biomarkers of interest, as exemplified by e.g. the effect of

proteases on GM-CSF.

It seems further likely that both cytokines and cell surface proteins can be cleaved. Thus, proteases could exert numerous effects on e.g. cell viability, differentiation and activation status. Importantly, in order to be able to test these and similar substances *in vitro*, it is important to carefully choose concentrations not exceeding a certain cytotoxicity. The responses may otherwise be dominated by events reflecting cytotoxicity; comparable to how the irritant properties of certain contact allergens can mask the effects linked to sensitization (Pedersen et al., 2004). On the other hand, irritant or cytotoxic effects may also allow or facilitate sensitization by creating a pro-inflammatory environment (“adjuvant effect”) as shown for several skin sensitizing chemicals (Pedersen et al., 2004; Grabbe et al., 1996; Martin, 2014). For proteases, impairment of barrier functions and thus higher permeability may play a role as well (Scheurer et al., 2015; Schweigert et al., 2000).

Taken together, our results indicate that the identified 391 potential biomarkers can be useful in order to predict the sensitizing ability of respiratory protein allergens, including proteases. Although the results of a series of cross-validations support a valid model (Fig. 3), the actual applicability of the biomarkers needs to be established by a set of external samples. The chosen signature seems to reflect relevant biological pathways. Next steps include challenging the model with further protein allergens and related non-allergenic control proteins. Especially, we intend to test food components comprising allergenic and non-allergenic proteins from different species.

Transparency document

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tiv.2017.09.029>.

Transparency document

The <http://dx.doi.org/10.1016/j.tiv.2017.09.029> associated with this article can be found, in online version.

Acknowledgements

This work was supported by grants from The ÅForsk Foundation (15-392) and the Faculty of Engineering (LTH). We would like to thank Ann-Charlott Olsson and Anna-Lovisa Nynäs for technical support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tiv.2017.09.029>.

References

- <http://imparas.eu/> 2015, accessed 161117.
http://www.serumsourceintl.com/questions_answers.html. 2015 [cited 2016 September 5th, 2016].
- Ather, J.L., et al., 2011. Airway epithelial NF-kappaB activation promotes allergic sensitization to an innocuous inhaled antigen. *Am. J. Respir. Cell Mol. Biol.* 44 (5), 631–638.
- Basketter, D.A., et al., 2010. Defining occupational and consumer exposure limits for enzyme protein respiratory allergens under REACH. *Toxicology* 268 (3), 165–170.
- Baur, X., 2005. Enzymes as occupational and environmental respiratory sensitizers. *Int. Arch. Occup. Environ. Health* 78 (4), 279–286.
- Baur, X., Budnik, L.T., von Kirchbach, G., 2013. Allergic asthma caused by exposure to bacterial alpha-amylase Termamyl(R). *Am. J. Ind. Med.* 56 (3), 378–380.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57 (1), 289–300.
- Boverhof, D.R., et al., 2008. Respiratory sensitization and allergy: current research approaches and needs. *Toxicol. Appl. Pharmacol.* 226 (1), 1–13.
- Budnik, L.T., et al., 2017. Sensitising effects of genetically modified enzymes used in flavour, fragrance, detergent and pharmaceutical production: cross-sectional study. *Occup. Environ. Med.* 74, 39–45.
- Carlsson, A., et al., 2011. Molecular serum portraits in patients with primary breast cancer predict the development of distant metastases. *Proc. Natl. Acad. Sci. U. S. A.* 108 (34), 14252–14257.
- Chapman, M.D., Wunschmann, S., Pomes, A., 2007. Proteases as Th2 adjuvants. *Curr. Allergy Asthma Rep.* 7 (5), 363–367.
- Chiou, Y.L., Lin, C.Y., 2009. Der p2 activates airway smooth muscle cells in a TLR2/MyD88-dependent manner to induce an inflammatory response. *J. Cell. Physiol.* 220 (2), 311–318.
- Corsini, E., et al., 2013. Role of oxidative stress in chemical allergens induced skin cells activation. *Food Chem. Toxicol.* 61, 74–81.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20 (3), 273–297.
- DeGroot, A., 2008. Patch Testing. Test Concentrations and Vehicles for 4350 Chemicals, 3rd edition. Acdegroot publishing.
- Forreryd, A., et al., 2015. Prediction of chemical respiratory sensitizers using GARD, a novel *in vitro* assay based on a genomic biomarker signature. *PLoS One* 10 (3), e0118808.
- Gough, L., et al., 1999. The cysteine protease activity of the major dust mite allergen Der p 1 selectively enhances the immunoglobulin E antibody response. *J. Exp. Med.* 190 (12), 1897–1902.
- Gough, L., et al., 2003. Proteolytic activity of the house dust mite allergen Der p 1 enhances allergenicity in a mouse inhalation model. *Clin. Exp. Allergy* 33 (8), 1159–1163.
- Grabbe, S., et al., 1996. Dissection of antigenic and irritative effects of epicutaneously applied haptens in mice. Evidence that not the antigenic component but nonspecific proinflammatory effects of haptens determine the concentration-dependent elicitation of allergic contact dermatitis. *J. Clin. Invest.* 98 (5), 1158–1164.
- Johansson, H., et al., 2011. A genomic biomarker signature can predict skin sensitizers using a cell-based *in vitro* alternative to animal tests. *BMC Genomics* 12, 399.
- Johansson, H., et al., 2013. The GARD assay for assessment of chemical skin sensitizers. *Toxicol. In Vitro* 27 (3), 1163–1169.
- Johansson, H., et al., 2014. Genomic allergen rapid detection in-house validation—a proof of concept. *Toxicol. Sci.* 139 (2), 362–370.
- Key Pathway Advisor by Clarivate Analytics (Formerly the IP & Science business of Thomson Reuters). <http://ipscience.thomsonreuters.com/product/metacore/>. 2016.
- King, C., et al., 1998. Dust mite proteolytic allergens induce cytokine release from cultured airway epithelium. *J. Immunol.* 161 (7), 3645–3651.
- König, K., et al., 2015. Cytokine profiles in nasal fluid of patients with seasonal or persistent allergic rhinitis. *Allergy, Asthma Clin. Immunol.* 11 (1), 26.
- Kool, M., et al., 2011. An unexpected role for uric acid as an inducer of T helper 2 cell immunity to inhaled antigens and inflammatory mediator of allergic asthma. *Immunity* 34 (4), 527–540.
- Luechtefeld, T., et al., 2016. Analysis of publically available skin sensitization data from REACH registrations 2008–2014. *ALTEX* 33 (2), 135–148.
- Martin, S.F., 2014. Adaptation in the innate immune system and heterologous innate immunity. *Cell. Mol. Life Sci.* 71 (21), 4115–4130.
- McClain, S., et al., 2014. Allergic sensitization: food- and protein-related factors. *Clin. Transl. Allergy* 4 (1), 11.
- Noble, W.S., 2006. What is a support vector machine? *Nat. Biotechnol.* 24 (12), 1565–1567.
- North, C.M., et al., 2016. Developing a framework for assessing chemical respiratory sensitization: a workshop report. *Regul. Toxicol. Pharmacol.* 80, 295–309.
- OECD, 2012. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. pp. 1–59.
- OECD, E.C.S.J.R.C., the United States Environmental Protection Agency, the US Army Engineer Research & Development Center, 2014. User's Handbook Supplement to the Guidance Document for Developing and Assessing AOPs.
- Pedersen, L.K., et al., 2004. Augmentation of skin response by exposure to a combination of allergens and irritants - a review. *Contact Dermatitis* 50 (5), 265–273.
- Piccolo, S.R., et al., 2012. A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics* 100 (6), 337–344.
- Piccolo, S.R., et al., 2013. Multiplatform single-sample estimates of transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* 110 (44), 17778–17783.
- Porter, P., et al., 2009. Link between allergic asthma and airway mucosal infection suggested by proteinase-secreting household fungi. *Mucosal Immunol.* 2 (6), 504–517.
- Roggen, E.L., 2014. *In vitro* approaches for detection of chemical sensitization. *Basic Clin. Pharmacol. Toxicol.* 115 (1), 32–40.
- Breakdown of the molecular processes driving the adverse outcome pathways (AOPs) of skin and respiratory sensitization induction in humans exposed to xenobiotics and proteins. In: Roggen, E.L., E.C.A.H.V., Loveren (Eds.), *Molecular Immunotoxicology*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
- Sarlo, K., et al., 1997. Proteolytic detergent enzymes enhance the allergic antibody responses of guinea pigs to nonproteolytic detergent enzymes in a mixture: implications for occupational exposure. *J. Allergy Clin. Immunol.* 100 (4), 480–487.
- Scheurer, S., Toda, M., Vieths, S., 2015. What makes an allergen? *Clin. Exp. Allergy* 45 (7), 1150–1161.
- Schuurhuis, D.H., et al., 2006. Ins and outs of dendritic cells. *Int. Arch. Allergy Immunol.* 140 (1), 53–72.
- Schweigert, M.K., Mackenzie, D.P., Sarlo, K., 2000. Occupational asthma and allergy associated with the use of enzymes in the detergent industry—a review of the epidemiology, toxicology and methods of prevention. *Clin. Exp. Allergy* 30 (11), 1511–1518.
- Straube, F., et al., 2005. Contact allergens and irritants show discrete differences in the activation of human monocyte-derived dendritic cells: consequences for *in vitro* detection of contact allergens. *Arch. Toxicol.* 79 (1), 37–46.
- Takai, T., et al., 2014. TSLP expression induced via Toll-like receptor pathways in human keratinocytes. *Methods Enzymol.* 535, 371–387.
- Thomas, W.R., et al., 2002. Characterization and immunobiology of house dust mite allergens. *Int. Arch. Allergy Immunol.* 129 (1), 1–18.
- Yazdi, A.S., Ghoreschi, K., Rocken, M., 2007. Inflammasome activation in delayed-type hypersensitivity reactions. *J. Invest. Dermatol.* 127 (8), 1853–1855.

Supplementary data

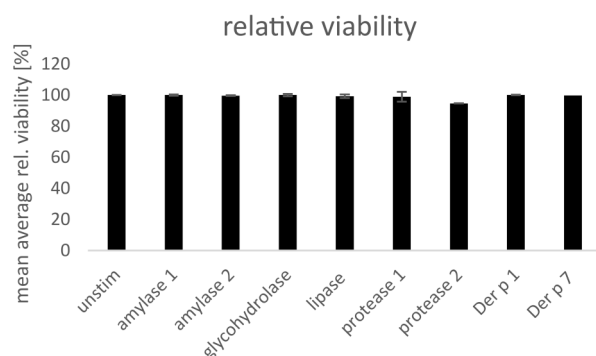


Figure S1 Viability assessed with propidium iodide staining and flow cytometry; cells were treated with the indicated substances for 24 hours (n=3, error bars show standard deviations).

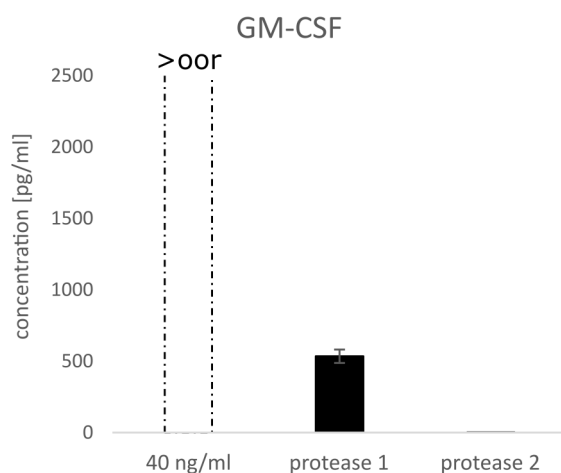


Figure S2 Changes in GM-CSF protein levels in MUTZ-3 supernatants after stimulation with the indicated substances for 24 hours, as determined by Bio-Plex® technology (n=3, error bars show standard deviations). GM-CSF was part of the complete cell culture medium (added to 40 ng/ml) in all samples, which is reflected by values out of range (oor) of the upper part of the standard curve (left bar), proteases seem to degrade GM-CSF (middle and right bar).

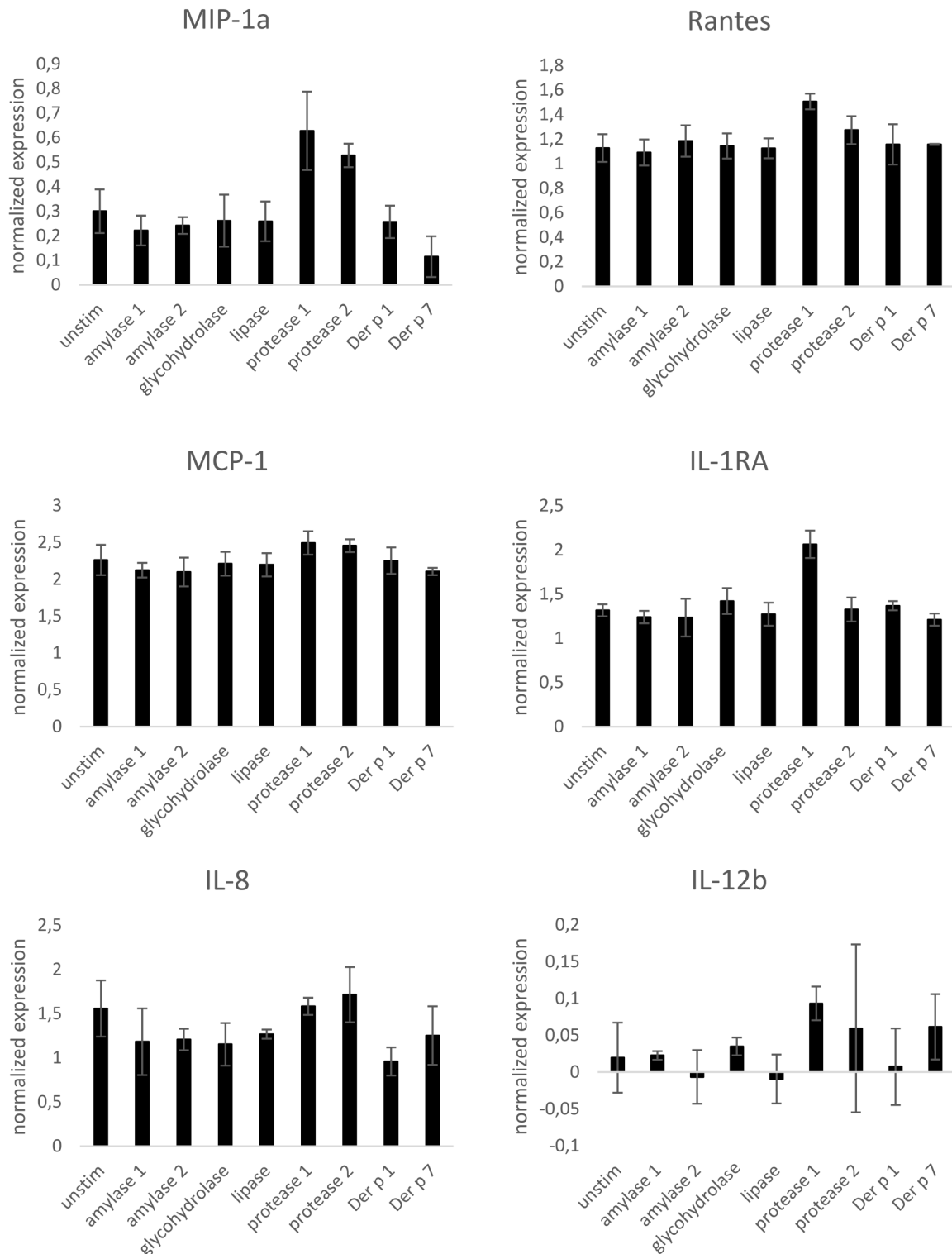


Figure S3 RNA microarray expression data of the indicated cytokines. (n=3, error bars show standard deviations. Microarray data is normalized using the scan algorithm (Piccolo et al., 2012; Piccolo et al., 2013), and should thus not be interpreted in a linear fashion).

Supplementary Table 1 391 transcripts – protein biomarker signature

Transcript Cluster ID	Gene Symbol	Gene Title
8083409	AADACL2	arylacetamide deacetylase-like 2
7897960	AADACL3	arylacetamide deacetylase-like 3
8137330	ABCB8	ATP binding cassette subfamily B member 8
8154563	ACER2	alkaline ceramidase 2
7936968	ADAM12	ADAM metallopeptidase domain 12
8164766	AK8	adenylate kinase 8
8087433	AMT /// NICN1	aminomethyltransferase /// nicotin 1
7938951	ANO5	anoctamin 5
8107421	AP3S1	adaptor-related protein complex 3, sigma 1 subunit
8007584	ASB16	ankyrin repeat and SOCS box containing 16
8141150	ASNS	asparagine synthetase (glutamine-hydrolyzing)
8052413	ATP1B3P1	ATPase, Na ⁺ /K ⁺ transporting, beta 3 pseudogene
7971731	ATP7B	ATPase, Cu ⁺⁺ transporting, beta polypeptide
8155026	ATP8B5P	ATPase, class I, type 8B, member 5, pseudogene
7915870	ATPAF1	ATP synthase mitochondrial F1 complex assembly factor 1
8073799	ATXN10	ataxin 10
7964660	AVPR1A	arginine vasopressin receptor 1A
7906205	BCAN	brevican
8160260	BNC2	basonuclin 2
8170159	BRS3	bombesin-like receptor 3
7933204	C10orf10	chromosome 10 open reading frame 10
7936996	C10orf90	chromosome 10 open reading frame 90
7955119	C12orf54	chromosome 12 open reading frame 54
8008980	C17orf82	chromosome 17 open reading frame 82
8021774	C18orf65	chromosome 18 open reading frame 65
8052698	C1D	C1D nuclear receptor corepressor
7934729	C1DP3	C1D nuclear receptor corepressor pseudogene 3
7934731	C1DP3 /// C1DP2	C1D nuclear receptor corepressor pseudogene 3 /// C1D nuclear receptor corepressor pseudogene 2
7912349	C1orf127	chromosome 1 open reading frame 127
8083223	C3orf58	chromosome 3 open reading frame 58

8103684	C4orf27	chromosome 4 open reading frame 27
8159078	CACFD1	calcium channel flower domain containing 1
8088491	CADPS	Ca ⁺⁺ -dependent secretion activator
8035146	CALR3	calreticulin 3
7906552	CASQ1	calsequestrin 1 (fast-twitch, skeletal muscle)
8135458	CBLL1	Cbl proto-oncogene-like 1, E3 ubiquitin protein ligase
8001329	CBLN1	cerebellin 1 precursor
8072153	CCDC117	coiled-coil domain containing 117
7958711	CCDC63	coiled-coil domain containing 63
8102643	CCNA2	cyclin A2
7904287	CD2	CD2 molecule
8019478	CD7	CD7 molecule
8142878	CDC26	cell division cycle 26
7982868	CHAC1	ChaC glutathione-specific gamma-glutamylcyclotransferase 1
8131140	CHST12 /// LOC101927181	carbohydrate (chondroitin 4) sulfotransferase 12 /// uncharacterized LOC101927181
8173524	CITED1	Cbp/p300-interacting transactivator, with Glu/Asp rich carboxy-terminal domain, 1
7992782	CLDN9	claudin 9
8083034	CLSTN2	calsyntenin 2
8135099	COL26A1	collagen, type XXVI, alpha 1
7957008	CPSF6	cleavage and polyadenylation specific factor 6
7921449	CRP	C-reactive protein, pentraxin-related
8095412	CSN1S1	casein alpha s1
8064351	CSNK2A1	casein kinase 2, alpha 1 polypeptide
8001666	CSNK2A2	casein kinase 2, alpha prime polypeptide
8095680	CXCL8	chemokine (C-X-C motif) ligand 8
8175524	CXorf66	chromosome X open reading frame 66
7988767	CYP19A1	cytochrome P450, family 19, subfamily A, polypeptide 1
7984862	CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2
8036969	CYP2A6 /// CYP2A13	cytochrome P450, family 2, subfamily A, polypeptide 6 /// cytochrome P450, family 2, subfamily A, polypeptide 13
7929478	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19
8155418	CYP4F30P /// FAM95B1 /// LOC105379252	cytochrome P450, family 4, subfamily F, polypeptide 30, pseudogene /// family with sequence similarity 95, member B1 /// coiled-coil domain-containing protein 29-like
8155627	CYP4F30P ///	cytochrome P450, family 4, subfamily F, polypeptide 30, pseudogene /// family with

	FAM95B1 /// LOC105379252	sequence similarity 95, member B1 /// coiled-coil domain-containing protein 29-like
8161381	CYP4F30P /// FAM95B1 /// LOC105379252	cytochrome P450, family 4, subfamily F, polypeptide 30, pseudogene /// family with sequence similarity 95, member B1 /// coiled-coil domain-containing protein 29-like
7974689	DACT1	dishevelled-binding antagonist of beta-catenin 1
7969959	DAOA	D-amino acid oxidase activator
7928308	DDIT4	DNA damage inducible transcript 4
7981326	DIO3OS /// MIR1247	DIO3 opposite strand/antisense RNA (head to head) /// microRNA 1247
8131719	DNAH11	dynein, axonemal, heavy chain 11
8004957	DNAH9	dynein, axonemal, heavy chain 9
7991512	DNM1P46	dynamamin 1 pseudogene 46
7920633	DPM3	dolichyl-phosphate mannosyltransferase polypeptide 3
7903079	DR1	down-regulator of transcription 1
8115831	DUSP1	dual specificity phosphatase 1
8116548	DUSP22	dual specificity phosphatase 22
8123819	EEF1E1 /// EEF1E1- BLOC1S5	eukaryotic translation elongation factor 1 epsilon 1 /// EEF1E1-BLOC1S5 readthrough (NMD candidate)
8108370	EGR1	early growth response 1
7939314	EHF	ets homologous factor
8127145	ELOVL5	ELOVL fatty acid elongase 5
8126750	ENPP5	ectonucleotide pyrophosphatase/phosphodiesterase 5 (putative)
7979179	ERO1A	endoplasmic reticulum oxidoreductase alpha
8056959	EVX2	even-skipped homeobox 2
8175299	FAM127C	family with sequence similarity 127, member C
8105607	FAM159B	family with sequence similarity 159, member B
8149324	FAM167A	family with sequence similarity 167, member A
8157144	FAM206A	family with sequence similarity 206, member A
8007548	FAM215A	family with sequence similarity 215, member A (non-protein coding)
7896756	FAM87B	family with sequence similarity 87, member B
7982574	FAM98B	family with sequence similarity 98, member B
7906475	FCRL6	Fc receptor-like 6
7979204	FERMT2	fermitin family member 2
8114805	FGF1	fibroblast growth factor 1 (acidic)

8139723	FKBP9P1	FK506 binding protein 9 pseudogene 1
7954692	FLJ13224	uncharacterized LOC79857
8065758	FLJ38773	uncharacterized protein FLJ38773
7975779	FOS	FBJ murine osteosarcoma viral oncogene homolog
8029693	FOSB	FBJ murine osteosarcoma viral oncogene homolog B
7915408	FOXJ3	forkhead box J3
8027770	FXYP7	FXYP domain containing ion transport regulator 7
7933075	FZD8	frizzled class receptor 8
8128087	GABRR1	gamma-aminobutyric acid (GABA) A receptor, rho 1
8155802	GDA	guanine deaminase
8165032	GLT6D1	glycosyltransferase 6 domain containing 1
7989476	GOLGA2P11	golgin A2 pseudogene 11
8083839	GPR160	G protein-coupled receptor 160
8044124	GPR45	G protein-coupled receptor 45
8021461	GRP	gastrin-releasing peptide
8066384	GTSF1L	gametocyte specific factor 1-like
8090091	HACD2	3-hydroxyacyl-CoA dehydratase 2
7906197	HAPLN2	hyaluronan and proteoglycan link protein 2
8180029	HLA-DQB2	major histocompatibility complex, class II, DQ beta 2
8169984	HPRT1	hypoxanthine phosphoribosyltransferase 1
8103722	HSP90AA6P	heat shock protein 90kDa alpha (cytosolic), class A member 6, pseudogene
8139592	HUS1 /// PKD1L1	HUS1 checkpoint clamp component /// polycystic kidney disease 1 like 1
7925846	IDI2-AS1	IDI2 antisense RNA 1
8023175	IER3IP1	immediate early response 3 interacting protein 1
8160383	IFNA14 /// IFNA7	interferon, alpha 14 /// interferon, alpha 7
8053715	IGKV2D-30 /// IGKV2D-29 /// IGKV1D-17 /// IGKV2D-18	immunoglobulin kappa variable 2D-30 /// immunoglobulin kappa variable 2D-29 /// immunoglobulin kappa variable 1D-17 /// immunoglobulin kappa variable 2D-18 (pseudogene)
8072004	IGLL3P	immunoglobulin lambda-like polypeptide 3, pseudogene
8014891	IKZF3	IKAROS family zinc finger 3
8044563	IL1F10	interleukin 1 family, member 10 (theta)
8080138	IQCF2	IQ motif containing F2
7969482	IRG1	immunoresponsive 1 homolog (mouse)

7995674	IRX6	iroquois homeobox 6
8032465	JSRP1	junctional sarcoplasmic reticulum protein 1
8028332	KCNK6	potassium channel, two pore domain subfamily K, member 6
8074106	KLHDC7B	kelch domain containing 7B
8015456	KLHL11	kelch-like family member 11
7977567	KLHL33	kelch-like family member 33
8030753	KLK3	kallikrein related peptidase 3
8038655	KLK4	kallikrein related peptidase 4
7961198	KLRAP1	killer cell lectin-like receptor subfamily A pseudogene 1
7961865	KRAS	Kirsten rat sarcoma viral oncogene homolog
7944952	KRT18P59	keratin 18 pseudogene 59
8015060	KRT24	keratin 24, type I
8156846	KRT8P11	keratin 8 pseudogene 11
8077499	LINC00312 /// LMCD1	long intergenic non-protein coding RNA 312 /// LIM and cysteine-rich domains 1
8076819	LINC00898	long intergenic non-protein coding RNA 898
8026503	LINC00905	long intergenic non-protein coding RNA 905
7961413	LINC01559	long intergenic non-protein coding RNA 1559
8050113	LOC100129581	uncharacterized LOC100129581
7981190	LOC100130815	uncharacterized LOC100130815
8071168	LOC101929738 /// LOC102724151	putative POM121-like protein 1 /// putative POM121-like protein 1
7986291	LOC105370980	putative uncharacterized protein UNQ9370/PRO34162
7995310	LOC162137	uncharacterized LOC162137
8130553	LOC401281 /// LOC105378085	uncharacterized LOC401281 /// uncharacterized LOC105378085
8110916	LOC442132	golgin A6 family-like 1 pseudogene
7950555	LRRC32	leucine rich repeat containing 32
7940002	LRRC55	leucine rich repeat containing 55
8089830	LRRC58	leucine rich repeat containing 58
8049530	LRRFIP1	leucine rich repeat (in FLII) interacting protein 1
8002969	MAF	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog
7997740	MAP1LC3B	microtubule-associated protein 1 light chain 3 beta
7925504	MAP1LC3C	microtubule associated protein 1 light chain 3 gamma
8103341	MAP9	microtubule-associated protein 9

8007828	MAPT	microtubule associated protein tau
8166382	MBTPS2	membrane bound transcription factor peptidase, site 2
7919751	MCL1	myeloid cell leukemia 1
7978801	MDGA2	MAM domain containing glycosylphosphatidylinositol anchor 2
8011968	MED31	mediator complex subunit 31
7985920	MESP2	mesoderm posterior bHLH transcription factor 2
7903294	MFSD14A	major facilitator superfamily domain containing 14A
8012535	MFSD6L	major facilitator superfamily domain containing 6-like
8136709	MGAM2	maltase-glucoamylase 2 (putative)
8010113	MGAT5B	mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase, isozyme B
8109159	MIR145 /// CARMN	microRNA 145 /// cardiac mesoderm enhancer-associated non-coding RNA
7971661	MIR16-1 /// MIR15A	microRNA 16-1 /// microRNA 15a
8008885	MIR21 /// VMP1	microRNA 21 /// vacuole membrane protein 1
8172266	MIR221	microRNA 221
8073544	MIR33A	microRNA 33a
7943736	MIR34B	microRNA 34b
8059026	MIR375	microRNA 375
8175256	MIR503HG	MIR503 host gene
8053064	MOB1A	MOB kinase activator 1A
7952046	MPZL2	myelin protein zero-like 2
8049375	MROH2A	maestro heat-like repeat family member 2A
7912863	MST1P2 /// MST1L /// MST1 /// LOC105369252 /// LOC102724562	macrophage stimulating 1 (hepatocyte growth factor-like) pseudogene 2 /// macrophage stimulating 1-like /// macrophage stimulating 1 /// hepatocyte growth factor-like protein /// hepatocyte growth factor-like protein
8104074	MTNR1A	melatonin receptor 1A
8012726	MYH1	myosin, heavy chain 1, skeletal muscle, adult
8089329	MYH15	myosin, heavy chain 15
8009873	MYO15B	myosin XVB
7943051	NAALAD2	N-acetylated alpha-linked acidic dipeptidase 2
8101723	NAP1L5	nucleosome assembly protein 1-like 5
7957298	NAV3	neuron navigator 3

8028600	NCCRP1	non-specific cytotoxic cell receptor protein 1 homolog (zebrafish)
8031387	NCR1	natural cytotoxicity triggering receptor 1
8001197	NETO2	neuropilin (NRP) and tolloid (TLL)-like 2
7986665	NIPA2	non imprinted in Prader-Willi/Angelman syndrome 2
8139367	NPC1L1	NPC1-like 1
8116607	NQO2 /// LINC01011	NAD(P)H dehydrogenase, quinone 2 /// long intergenic non-protein coding RNA 1011
8055952	NR4A2	nuclear receptor subfamily 4, group A, member 2
7965573	NTN4	netrin 4
8021357	ONECUT2	one cut homeobox 2
8120151	OPN5	opsin 5
7962792	OR10AD1	olfactory receptor, family 10, subfamily AD, member 1
7921356	OR10K2	olfactory receptor, family 10, subfamily K, member 2
7937975	OR51A8P	olfactory receptor, family 51, subfamily A, member 8 pseudogene
7937971	OR51N1P	olfactory receptor, family 51, subfamily N, member 1 pseudogene
7946021	OR52A4P	olfactory receptor, family 52, subfamily A, member 4, pseudogene
8088903	OR7E121P	olfactory receptor, family 7, subfamily E, member 121 pseudogene
8142019	ORC5	origin recognition complex subunit 5
8105040	OSMR	oncostatin M receptor
7965884	PAH	phenylalanine hydroxylase
8057004	PDE11A	phosphodiesterase 11A
8080676	PDE12	phosphodiesterase 12
7951165	PGR	progesterone receptor
7961604	PLCZ1	phospholipase C, zeta 1
8091306	PLSCR4	phospholipid scramblase 4
8165575	PNPLA7	patatin-like phospholipase domain containing 7
7907024	POGK	pogo transposable element with KRAB domain
8106820	POLR3G	polymerase (RNA) III (DNA directed) polypeptide G (32kD)
8074714	POM121L7 /// LOC102724728 /// POM121L8P /// LOC101929738 /// LOC102724151	POM121 transmembrane nucleoporin-like 7 /// putative POM121-like protein 1 /// POM121 transmembrane nucleoporin-like 8 pseudogene /// putative POM121-like protein 1 /// putative POM121-like protein 1
8100990	PPBPP2	pro-platelet basic protein pseudogene 2
7923659	PPP1R15B	protein phosphatase 1, regulatory subunit 15B

7897991	PRAMEF8 /// PRAMEF7	PRAME family member 8 /// PRAME family member 7
7912591	PRAMEF8 /// PRAMEF7	PRAME family member 8 /// PRAME family member 7
7912606	PRAMEF8 /// PRAMEF7	PRAME family member 8 /// PRAME family member 7
8176806	PRY2 /// PRY /// PRYP3 /// PRYP4	PTPN13-like, Y-linked 2 /// PTPN13-like, Y-linked /// PTPN13-like, Y-linked pseudogene 3 /// PTPN13-like, Y-linked pseudogene 4
8176935	PRY2 /// PRY /// PRYP3 /// PRYP4	PTPN13-like, Y-linked 2 /// PTPN13-like, Y-linked /// PTPN13-like, Y-linked pseudogene 3 /// PTPN13-like, Y-linked pseudogene 4
8177323	PRY2 /// PRY /// PRYP3 /// PRYP4	PTPN13-like, Y-linked 2 /// PTPN13-like, Y-linked /// PTPN13-like, Y-linked pseudogene 3 /// PTPN13-like, Y-linked pseudogene 4
8177395	PRY2 /// PRY /// PRYP3 /// PRYP4	PTPN13-like, Y-linked 2 /// PTPN13-like, Y-linked /// PTPN13-like, Y-linked pseudogene 3 /// PTPN13-like, Y-linked pseudogene 4
7935990	PSD	pleckstrin and Sec7 domain containing
7922976	PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)
8160040	PTPRD	protein tyrosine phosphatase, receptor type, D
8080781	PXK	PX domain containing serine/threonine kinase
8040618	RAB10	RAB10, member RAS oncogene family
8146564	RAB2A	RAB2A, member RAS oncogene family
8141843	RASA4 /// LOC102724229	RAS p21 protein activator 4 /// uncharacterized LOC102724229
7977003	RCOR1	REST corepressor 1
8161192	RNF38	ring finger protein 38
7916882	RPE65	retinal pigment epithelium-specific protein 65kDa
7938295	RPL27A	ribosomal protein L27a
8092654	RPL39L	ribosomal protein L39-like
8055941	RPRM	reprimin, TP53 dependent G2 arrest mediator candidate
8021468	RPS26P54	ribosomal protein S26 pseudogene 54
8092686	RTP2	receptor (chemosensory) transporter protein 2
8104781	RXFP3	relaxin/insulin-like family peptide receptor 3
7920264	S100A5	S100 calcium binding protein A5
7940622	SCGB1D1	secretoglobin, family 1D, member 1
8046020	SCN2A	sodium channel, voltage gated, type II alpha subunit

8108099	SEC24A	SEC24 homolog A, COPII coat complex component
8071274	SEPT5-GP1BB	SEPT5-GP1BB readthrough
7934698	SFTPA2 /// SFTPA1	surfactant protein A2 /// surfactant protein A1
7934708	SFTPA2 /// SFTPA1	surfactant protein A2 /// surfactant protein A1
8149438	SGCZ	sarcoglycan zeta
7972461	SLC15A1	solute carrier family 15 (oligopeptide transporter), member 1
7957819	SLC17A8	solute carrier family 17 (vesicular glutamate transporter), member 8
8154135	SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1
8107909	SLC22A4	solute carrier family 22 (organic cation/zwitterion transporter), member 4
8108199	SLC25A48	solute carrier family 25, member 48
8135436	SLC26A4	solute carrier family 26 (anion exchanger), member 4
8141922	SLC26A5	solute carrier family 26 (anion exchanger), member 5
7951485	SLC35F2	solute carrier family 35, member F2
8111417	SLC45A2	solute carrier family 45, member 2
8173414	SLC7A3	solute carrier family 7 (cationic amino acid transporter, y+ system), member 3
8015037	SMARCE1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1
7981998	SNORD116-25	small nucleolar RNA, C/D box 116-25
7982000	SNORD116-26	small nucleolar RNA, C/D box 116-26
8114468	SNORD63	small nucleolar RNA, C/D box 63
7952797	SPATA19	spermatogenesis associated 19
8028389	SPRED3	sprouty-related, EVH1 domain containing 3
7971644	SPRYD7	SPRY domain containing 7
8150722	ST18	suppression of tumorigenicity 18, zinc finger
7986229	ST8SIA2	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2
8057771	STAT4	signal transducer and activator of transcription 4
8000590	SULT1A1	sulfotransferase family 1A member 1
8122684	SUMO4	small ubiquitin-like modifier 4
7947147	SVIP	small VCP/p97-interacting protein
8116874	SYCP2L	synaptonemal complex protein 2-like
7951038	TAF1D /// SNORA40	TATA box binding protein associated factor 1D /// small nucleolar RNA, H/ACA box 40
8045889	TANC1	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1

8127932	TBX18	T-box 18
8136557	TBXAS1	thromboxane A synthase 1 (platelet)
8146914	TERF1	telomeric repeat binding factor (NIMA-interacting) 1
8054364	TGFBRAP1	transforming growth factor beta receptor associated protein 1
8097704	TMEM184C	transmembrane protein 184C
8094830	TMEM33	transmembrane protein 33
8157092	TMEM38B	transmembrane protein 38B
8043100	TMSB10	thymosin beta 10
8149749	TNFRSF10D	tumor necrosis factor receptor superfamily, member 10d, decoy with truncated death domain
8033248	TNFSF14	tumor necrosis factor (ligand) superfamily, member 14
7995267	TP53TG3D /// LOC102723655 /// LOC102723713	TP53 target 3D /// TP53-target gene 3 protein /// TP53-target gene 3 protein
7971375	TPT1	tumor protein, translationally-controlled 1
8104107	TRIML2	tripartite motif family like 2
8161774	TRPM6	transient receptor potential cation channel, subfamily M, member 6
8065603	TSPY26P	testis specific protein, Y-linked 26, pseudogene
8129097	TSPYL1	TSPY-like 1
8075142	TTC28	tetratricopeptide repeat domain 28
8176692	TTY9B /// OFD1P6Y /// OFD1P5Y /// TTY9A	testis-specific transcript, Y-linked 9B (non-protein coding) /// OFD1 pseudogene 6, Y-linked /// OFD1 pseudogene 5, Y-linked /// testis-specific transcript, Y-linked 9A (non-protein coding)
8177195	TTY9B /// OFD1P6Y /// OFD1P5Y /// TTY9A	testis-specific transcript, Y-linked 9B (non-protein coding) /// OFD1 pseudogene 6, Y-linked /// OFD1 pseudogene 5, Y-linked /// testis-specific transcript, Y-linked 9A (non-protein coding)
8094704	UBE2K	ubiquitin conjugating enzyme E2K
8157216	UGCG	UDP-glucose ceramide glucosyltransferase
8090366	UROC1	urocanate hydratase 1
8020842	WBP11P1	WW domain binding protein 11 pseudogene 1
8169022	WBP5	WW domain binding protein 5
8157818	WDR38	WD repeat domain 38
8144774	VPS37A	vacuolar protein sorting 37 homolog A (S. cerevisiae)
8140398	YWHAG	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma
8137219	ZBED6CL	ZBED6 C-terminal like

8036291	ZNF565	zinc finger protein 565
8036430	ZNF781 /// ZFP30	zinc finger protein 781 /// ZFP30 zinc finger protein
8035789	ZNF826P /// LOC101929991 /// LOC102724436	zinc finger protein 826, pseudogene /// zinc finger protein 93-like /// putative zinc finger protein 66
8039605	ZNF835	zinc finger protein 835
7902685	---	---
7904417	---	---
7905077	---	---
7914992	---	---
7919572	---	---
7924495	---	---
7925087	---	---
7927153	---	---
7932964	---	---
7934883	---	---
7941608	---	---
7942809	---	---
7942814	---	---
7943954	---	---
7944867	---	---
7945130	---	---
7951701	---	---
7953747	---	---
7957495	---	---
7959012	---	---
7960861	---	---
7973054	---	---
7973444	---	---
7974562	---	---
7976057	---	---
7980003	---	---
7981773	---	---
7998629	---	---

7999406	---	---
8005757	---	---
8010622	---	---
8013517	---	---
8017827	---	---
8018377	---	---
8035956	---	---
8036300	---	---
8036403	---	---
8046790	---	---
8053311	---	---
8054939	---	---
8055265	---	---
8055348	---	---
8055424	---	---
8055492	---	---
8058145	---	---
8067965	---	---
8068046	---	---
8070930	---	---
8076298	---	---
8084878	---	---
8084887	---	---
8091097	---	---
8091118	---	---
8091186	---	---
8091676	---	---
8095161	---	---
8095331	---	---
8098439	---	---
8099130	---	---
8099362	---	---
8099713	---	---
8102781	---	---

8110104	---	---
8111118	---	---
8112666	---	---
8113276	---	---
8113352	---	---
8121064	---	---
8122732	---	---
8124634	---	---
8127658	---	---
8129313	---	---
8131301	---	---
8137925	---	---
8138920	---	---
8143708	---	---
8143710	---	---
8143747	---	---
8146857	---	---
8155569	---	---
8161426	---	---
8161437	---	---
8163015	---	---
8166945	---	---
8167953	---	---
8175589	---	---
8180232	---	---
8180281	---	---
8180344	---	---

--- no assignments available