Tri-culture system for pro-hapten sensitizer identification and potency classification

Serom Lee[#1], **Talia Greenstein**[#1], **Lingting Shi**[1], **Tim Maguire**[1], **Rene Schloss**[1], and **Martin Yarmush**[1,2]

[1]Department of Biomedical Engineering, Rutgers, the State University of New Jersey, 599 Taylor Road, Piscataway, NJ 08854, USA

[2]Center for Engineering in Medicine and the Department of Surgery, Massachusetts General Hospital and the Shriners Burns Hospital, Boston, MA 02114, USA

[#] These authors contributed equally to this work.

Abstract

Allergic contact dermatitis (ACD) is an inflammatory disease that impacts 15-20% of the general population and accurate screening methods for chemical risk assessment are needed. However, most approaches poorly predict pre- and pro-hapten sensitizers, which require abiotic or metabolic conversion prior to inducing sensitization. We developed a tri-culture system comprised of MUTZ-3-derived Langerhans cells, HaCaT keratinocytes, and primary dermal fibroblasts to mimic the cellular and metabolic environments of skin sensitization. A panel of non-sensitizers and sensitizers was tested and the secretome was evaluated. A support vector machine (SVM) was used to identify the most predictive sensitization signature and classification trees identified statistical thresholds to predict sensitizer potency. The SVM computed 91% tri-culture prediction accuracy using the top 3 ranking biomarkers (IL-8, MIP-1β, and GM-CSF) and improved the detection of pre- and pro-haptens. This *in vitro* assay combined with *in silico* data analysis presents a promising approach and offers the possibility of multi-metric analysis for enhanced ACD sensitizer screening.

Keywords

IL-8; Tri-culture; Skin Sensitizer; Pre- and Pro-hapten; Classification Tree; Support Vector Machine

INNOVATION OF THE TECHNOLOGY

Predictive, *in vitro* sensitizer screening assays are critically important for biopharmaceutical industries, which constantly develop novel chemical formulations. However, current *in vitro* assays have particularly low accuracy in predicting pre- and pro-hapten sensitizers, which require chemical transformation or metabolic activation prior to inducing sensitization. Our innovative tri-culture system, designed to mimic the cellular and metabolic environments of

Correspondence should be addressed to R.S. (Schloss@soe.rutgers.edu).

skin sensitization, combined with *in silico* data analysis, offers the possibility of a high-accuracy, cost-effective, high-throughput, multi-metric analysis for enhanced skin sensitizer screening and potency assessment.

INTRODUCTION*

Allergic Contact Dermatitis (ACD) is one of the most prevalent dermatoses, impacting approximately 15–20% of the general population[1] and characterized by erythematous reactions generally located where the allergen was applied[2]. There are three classes of contact allergens, grouped according to the mechanistic pathways through which they form macro-molecular immunogens that initiate an allergic response. Haptens bind readily to skin proteins to form a sensitizing entity[3]. Pre- and pro-haptens require abiotic or metabolic activation prior to inducing an allergic response[4,5].

ACD poses a significant safety and occupational hazard, and therefore potentially sensitizing agents must be screened and classified. Until recently, animal testing was the gold standard used to screen and identify skin sensitizers, but high-associated costs, reduced accuracy relative to human clinical data, and the global push to ban animal testing of cosmetic products and ingredients have motivated research of alternative methods to screen drugs, cosmetics, and other chemical moieties[6]. For the accurate identification of pro-hapten sensitizers, a metabolic component must be incorporated into the screening tool.

Several strategies to introduce a metabolic component to transform pre- and pro-hapten sensitizers using cytochrome p450 enzyme cocktails, liver microsomes, and keratinocyte cell lines were previously investigated[7–9]. Keratinocytes (KCs) in the epidermis are primarily responsible for carrying out xenobiotic metabolism in the skin using phase I enzymes such as CYP1A1, important in generating reactive intermediates from pro-haptens[10–15]. KCs also produce several inflammatory mediators during ACD, such as members of the IL-1 family, TNF-α, IL-6, IL-18, and GM-CSF, which orchestrate various events during the sensitization phase[16–18]. Fibroblasts in the dermis also play an important sensitization role by secreting chemokines such as CXCL12 and cytokines such as TNF-α[19,20]. Fibroblast secreted factors may also regulate the expression of cytochrome p450 enzymes in keratinocytes[21]. Thus, an approach that includes both skin cell types in co-culture with Langerhans cells (LCs), the local immune response initiator cells, may confer additional benefits compared with keratinocytes alone.

Previously, we established a full-thickness skin model co-cultured with LCs to be used in concert with computational support vector machine analysis to identify a panel of predictive biomarkers for the classification of skin sensitizers[22]. Given the high cost and low through-putness of full-thickness skin models, the current studies were designed to determine the

---

efficacy of a simpler tri-culture approach, which preserves the benefits of keratinocyte and fibroblasts, especially for pre- and pro-sensitizer classification. In addition, classification tree learning[23] was used to identify statistical thresholds to distinguish sensitizers and non-sensitizers and classify sensitizer potency. Our results demonstrated the efficacy of our tri-culture model in identifying all sensitizer types, and the value of utilizing classification models to evaluate large *in vitro* data sets generated from the secretome.

MATERIALS AND METHODS

Cell culture

The HaCaT keratinocyte (KC) cell line and human primary dermal fibroblasts (FB) were maintained in DMEM (Gibco) with 10% fetal bovine serum (FBS) and 100 U/mL penicillin, 100 μg/mL streptomycin supplementation (i.e. 1% penicillin-streptomycin) at 37°C and 5% CO_2. Media was changed every 2–3 days until confluence. HaCaT cell line was gifted from NJ Center for Dermal Research and primary fibroblasts were gifted from Dr. Francois Berthiaume.

The 5637 human bladder carcinoma line was purchased from ATCC (Manassas, VA) and cultured as previously described[22]. Conditioned medium from 5637 culture was supplemented into the MUTZ-3 culture medium as per the guidelines from DSMZ.

The MUTZ-3 cell line was a donation from Massachusetts General Hospital (Boston, MA) and is available for purchase from DSMZ (Brauncshweig, Germany) and cultured and differentiated as previously described[22]. Differentiated MUTZ-LCs utilized in experiments were maintained in Maturation Medium (complete media without conditioned media supplement).

Chemicals and reagents

Test chemicals included both non-sensitizers and known skin sensitizers of every class and potency as classified by the LLNA (Table 1). All chemicals were purchased through Sigma-Aldrich and prepared in dimethyl-sulfoxide (DMSO) at a final concentration of ~0.1%. To determine dose response, three concentrations of each chemical were tested with the exception of SDS, DNCB, and VL, due to cytotoxicity. Concentration ranges were based on values commonly reported in the literature and for inclusion were required to be at least 50% viable by Alamar Blue analysis (methods described below).

Tri-culture of HaCaT keratinocytes, dermal fibroblasts, and MUTZ-3 Langerhans cells

HaCaT KCs and human dermal FBs were plated at 1.25×10^4 cells (each) per well in 96-well plates in complete DMEM the night before the start of the experiment (i.e. day 6 of MUTZ-LC differentiation). On day 7, after the KCs and FBs became adherent, the wells were washed with maturation medium and fully differentiated 2.5×10^4 MUTZ-LCs were added to each well in 220 μL of vehicle or chemical treatment diluted in maturation media in triplicate. The MUTZ-LCs were also plated as a mono-culture in parallel. Cells incubated for 48 hours at 37°C and 5% CO_2 and then viability was assessed using Alamar Blue

analysis and supernatants were collected and stored at −20°C for future ELISA and multiplex analyses (Fig. 1).

Viability

Viability of both culture platforms was analyzed using the Alamar Blue™ assay, performed according to the manufacturer's protocol and measured with a DTX80 multimode detector (Beckman-Coulter). The final time point used for analysis was 4 hours after addition of Alamar Blue. Viability of each condition was computed as follows:

$$Viability = \frac{Average\ absorbance\ of\ treatment}{Average\ absorbance\ of\ vehicle}$$

Cytokine secretion

Supernatant collected from both tri-culture and mono-culture systems treated with sensitizers and non-sensitizers were analyzed for IL-8 secretion using ELISA (Biolegend, San Diego, CA), following the manufacturer's instructions. The supernatants were also analyzed for 27 human cytokines (IL-1β, IL-1ra, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12, IL-13, IL-15, IL-17, Eotaxin, Basic FGF, G-CSF, GM-CSF, IFN-γ, IP-10, MCAF, MIP-1α, MIP-β, PDGF-BB, RANTES, TNF-α, and VEGF) using a Bioplex assay following the manufacturer's instructions (Bio-plex Human Cytokine 27-plex panel; Bio-Rad Laboratories, Hercules, CA, USA). For bio-plex analysis triplicate wells from each experiment were pooled and cytokine secretion for one concentration of each sensitizer and non-sensitizer from our panel was measured.

Data acquisition and analysis

A total of four independent experiments were conducted where each condition was tested in triplicate per experiment and used for subsequent data analysis. Raw secretion and viability were averaged across triplicates. Stimulation index (SI) of cytokine production from both the tri-culture and mono-culture systems was determined by normalizing the raw cytokine concentration to the condition viability and corresponding vehicle with the following equation:

$$Stimulation\ Index = \frac{\left(\frac{Raw\ secretion\ of\ condition}{Raw\ secretion\ of\ vehicle}\right)}{Viability\ of\ condition}$$

Statistical significance of IL-8 secretion measured by ELISA was determined at p 0.05 using ANOVA and Fisher's least significant difference (LSD) post-hoc analysis in Kaleidagraph. A machine learning method using a previously developed support vector machine (SVM) in MATLAB® was used to compute and rank the margin distance for each cytokine measured in the bio-plex assay[22]. We trained and tested the SVM with leave-one out cross-validation; 90% of the dataset was used to train and 10% to test. To reduce bias, training and test data were randomly chosen through 100 iterative runs of the SVM. A classification model of the most predictive metrics was identified by k-fold cross validation for each individual biomarker analyzed by the SVM and accuracy, sensitivity, and specificity

was computed. Classification trees were generated with the IL-8 secretion data using "ctree" (conditional interference trees) from the package "party" in R(3.2.2). This machine learning method utilizes recursive partitioning by conditional interference to identify the best binary split based on standardized linear statistics. Forty-eight hours after chemical treatment initiation, IL-8 SI was calculated and four data points from every chemical concentration were averaged. The "ctree" function from the "party" package in R was used to identify a SI threshold to distinguish between sensitizer (S) and non-sensitizer (NS) such that where y represents the average of four IL-8 SI data points from a chemical (*x*) concentration:

$$if \begin{cases} y \leq threshold, \ then \ x = NS \\ y > threshold, \ then \ x = S \end{cases}$$

The resulting decision trees function to determine the prediction accuracy of the two *in vitro* systems based on the thresholds identified. This algorithm further allows for potency classification and calculates the accuracy of potency prediction.

RESULTS

Evaluation of IL-8 in tri-culture with HaCaT KCs, dermal FBs, and MUTZ-LCs

Increased IL-8 secretion has been used by several investigators as an effective sensitization metric. Therefore, IL-8 secretion was initially used to characterize our tri-culture model's ability to distinguish between sensitizers and non-sensitizers. The tri-culture and the MUTZ-LCs alone were treated with chemical non-sensitizers and sensitizers and supernatants were analyzed by ELISA. The tri-culture system accurately identified at least one concentration for every sensitizer to have significantly increased IL-8 secretion, while the MUTZ-LCs failed to accurately identify a number of sensitizers, including the hapten CLD and pre-/pro-haptens CA, PPD, HQ, and pBQ (Supplementary Fig. 1). A representative panel of measured IL-8 secretion for a non-sensitizer and sensitizers of each potency demonstrates the ability of the tri-culture system to distinguish between sensitizers and non-sensitizers (Fig. 2). To understand which cell population is most involved in the pro-inflammatory response, we treated MUTZ-LCs, KCs, and FBs alone and in every co-culture combination with DNCB, EU, XYL, or SDS and analyzed IL-8 secretion by ELISA. We found that the IL-8 contribution from KCs and FBs was insignificant; the presence of MUTZ-LCs is necessary for significant IL-8 secretion (data not shown).

Support vector machine analysis of secretome data

In order to further probe the efficacy of our tri-culture, to compare responses of the tri-cultures with responses of MUTZ-3 cells alone, and to select and quantify additional metrics for sensitizer prediction, the supernatants of chemically treated cultures were analyzed using a 27-plex inflammatory Bio-Plex bead-based array. The support vector machine (SVM) was used to calculate the margin distance of separation between two classes of chemicals: non-sensitizer (negative) treatments and sensitizer (positive) treatments. Greater margin distances indicate a greater degree of separation between the two classes for any given metric. We used this information to rank each cytokine and build a classification model by selecting the

key features necessary for accurate prediction. The SVM also calculated the sensitivity and specificity of each classification model.

The SVM top 12 secretome biomarkers collected from the MUTZ-3 culture in ranked order were found to be IL-8, MIP-1β, IL-9, IL-17, MIP-1α, IL-1β, IL-15, RANTES, GM-CSF, MCP-1, IL-7, and Eotaxin, achieving the maximal accuracy of 87.2% accuracy in combination, specificity of 95.5% and sensitivity of 79.6%. When only IL-8 and MIP-1β were used in combination, an accuracy of 86.2% was achieved. The algorithm was unable to converge to compute accuracy for more than 12 metrics. In contrast, the top 12 secretome biomarkers collected from the tri-culture system in ranked order were IL-8, MIP-1β, GM-CSF, RANTES, IL-15, MCP-1, MIP-1a, IL-17, VEGF, IL-1β, G-CSF, and IL-13, achieving 86.7% accuracy in combination. The top three cytokines, IL-8, MIP-1β, and GM-CSF, offer a classification model with the highest achievable accuracy of 91.1% when used in combination with 92.7% sensitivity and 89.8% specificity (Table 2). Therefore, our SVM analysis identified unique sensitizer signatures for the mono and tri-culture systems. When examining all hapten classes the tri-culture system achieved better sensitizer classification results.

Next we compared the ability of our two culture systems to identify pro-haptens, which are often difficult to classify using most *in vitro* methods. The SVM was used to build a classification model to predict pre-/pro-hapten sensitization potential and determine the accuracy of pre-/pro-hapten prediction for both the mono-culture and tri-culture systems. IL-8 and MIP-1β achieved 87.2% accuracy in combination for the MUTZ-LC mono-culture system with 93.3% sensitivity and 80.5% specificity, while IL-8, RANTES, and GM-CSF achieved 90.2% accuracy for the tri-culture system with 92.7% sensitivity and 87.8% specificity. Therefore, our analysis indicated that the tri-culture could identify all hapten classes and that multi-metric analysis could improve classification accuracy.

IL-8 secretion to evaluate sensitizer potency

The basic SVM is a binary classifier that can process and rank multiple data sets (i.e. biomarkers). Classification trees allow for multiclass classification using one data set. As it has been suggested that IL-8 secretion could be used for potency prediction[24] and IL-8 was shown to be the most responsive cytokine by the SVM, IL-8 levels were measured to evaluate sensitizer potency using R classification trees. 3 weak, 4 moderate, 3 strong, and 2 extreme sensitizers (based on LLNA potency classification) were tested with the MUTZ-LC and tri-culture systems. The IL-8 SI data from the weak and moderate sensitizers were grouped together (W/M) and the data from the strong and extreme sensitizers were grouped together (S/E) and classification trees were generated in R to identify statistical thresholds (thresholds A and B) dividing the groups such that, when IL-8 SI y is input for chemical x:

$$if \begin{cases} y \leq threshold\ A, & then\ x = NS \\ threshold\ A < y \leq threshold\ B, & then\ x = W/M \\ y > threshold\ B, & then\ x = S/E \end{cases}$$

For every chemical concentration, 4 individual data points were averaged. 18 averaged non-sensitizer data points, 21 averaged W/M sensitizer data points, and 14 averaged S/E sensitizer data points were assessed using the classification algorithm based on "ctree" from the R package "party" If statistically significant break points exist, the classification tree will generate two nodes. The first node divides between all sensitizers and non-sensitizers. The second node generates from the sensitizer branch to divide W/M and S/E sensitizers. The nodes represent the SI thresholds for chemical classification. The classification tree for IL-8 SI in MUTZ-LCs alone identified a threshold of 1.366 SI with a p value of < 0.001 to distinguish sensitizer from non-sensitizer, and a threshold of 3.274 SI to distinguish between W/M and S/E sensitizers (Fig. 3). Based on these thresholds, the MUTZ-LCs alone misclassify CA at 1000 μM, CLD at 125 and 250 μM, IE at 1000 μM, MMP at 150 and 300 μM, and HQ at 100 μM as non-sensitizers, SA at 2000 μM as a W/M sensitizer, EU at 500 and 1000 μM as S/E, and HQ at 25 μM, pBQ at 10 and 25 μM, and PPD as W/M sensitizers, resulting in an overall accuracy of 73.6% (Table 3).

The classification tree for IL-8 SI in tri-culture identified a threshold of 1.743 SI with a p value < 0.001 to distinguish sensitizer from non-sensitizer, and a threshold of 3.963 SI to distinguish between W/M and S/E sensitizers (Fig. 4). Based on these thresholds, the tri-culture misclassifies SA at 2000 μM and VL as sensitizers and PPD at 62.5 and 125 μM and pBQ as W/M sensitizers, resulting in an overall accuracy of 83%. Based on the computed threshold, the tri-culture system did not predict any false negatives (Table 3). Collectively, our results demonstrate that the tri-culture system achieves greater accuracy in distinguishing pre- and pro-hapten sensitizers from non-sensitizers and offers a new method for potency prediction that focuses on inflammatory response to sensitizers.

DISCUSSION

Predictive, *in vitro* sensitizer screening assays are critically important for biopharmaceutical industries, which constantly develop novel chemical formulations. However, current *in vitro* assays have particularly low accuracy in predicting pre- and pro-hapten sensitizers, which require chemical transformation or metabolic activation prior to inducing sensitization. We previously developed a full-thickness skin model that showed promise in predicting sensitization by pre- and pro-haptens, but this model was expensive and low-throughput[22]. The current studies were designed to develop a system containing the relevant metabolizing and responsive skin components, in conjunction with computational tools, to identify and ultimately predict sensitization potential of pre and pro-haptens in a high-throughput, cost-effective manner.

Several strategies have previously been attempted to fill this void. These include using a skin-like cocktail of CYP enzymes (CYP1A1, CYP1B1, CYP2B6, CYP2E1, and CYP3A5)[9], which were enriched 1000-fold and may not be fully representative of the *in vivo* content and activity of these enzymes[24]. Human liver microsomes have been used as another alternative source of metabolism[7]. As liver microsomes are already extensively used by the pharmaceutical industry to perform drug metabolism and pharmacokinetic studies, these microsomes potentially offer a viable source of metabolism that is both cost effective and amenable for high-throughput screens. However, there are tissue-specific discrepancies

in cytochrome p450 enzymes[10,11]. Thus, a more physiologically relevant source of enzymes that are involved during sensitization, such as keratinocytes and fibroblast cell cultures, may be a more suitable, cost effective alternative. Full-thickness skin equivalents have shown promise as a model of skin for *in vitro* sensitization studies but are expensive and low-throughput to detect pro-haptens. This motivated our use of a culture model incorporating both skin cells and MUTZ-LCs to mimic the environment of a 3D full-thickness skin model in a 2D co-culture platform that could be implemented in 96-well plates.

IL-8 secretion was chosen as a starting metric of sensitization based on literature precedence[9,25–27]. IL-8 is a chemotactic factor produced by a wide variety of cells[28,29]. Furthermore, delayed-type hypersensitivity has been shown to depend on IL-8[30]. As such, it has been extensively studied as a biomarker of skin sensitization. Enhanced IL-8 mRNA expression in moDCs, THP-1s, and MUTZ-LCs and secretion in moDCs and MUTZ-LCs in response to sensitizers have been previously quantified[9,25–27,31,32]. The precedence and success of using IL-8 as a metric of predicting sensitization established its promise as an initial biomarker for our tri-culture system.

Although IL-8 is a suitable screening metric, there is a consensus that a single biomarker or assay is unlikely to be predictive of all skin sensitizers [27,33,34]. Thus, a tiered strategy was developed that evaluates several different metrics of sensitization together to make an informed prediction. While we have identified several predictive metrics using our limited chemical panel, we recognize that as the chemical panel is expanded to validate our *in vitro* approach and identify molecular signatures, it is feasible that a different metric panel will emerge. This was the case for the GARD assay, that initially utilized a 10-gene signature for skin sensitizer prediction[35]. However, when this method was adapted by industry and more chemicals were evaluated, the 10 gene signature was expanded to 200 genes. It is possible that a similar expansion of secreted proteins or other cellular metrics will be necessary as more chemicals are evaluated using our tri-culture approach.

Nevertheless, in these initial studies, we expanded our analysis from IL-8 alone to a multi-plex screen of cytokines, chemokines, and growth factors and a support vector machine was used to build a classification model with the highest achievable accuracy, sensitivity, and specificity of prediction. The SVM was used to rank margin distances that maximized the difference between sensitizer and non-sensitizer for every biomarker. Combinations of biomarkers were selected and accuracy of the resulting classification model was predicted. The tri-culture system achieved its highest accuracy of 91% with the top 3 biomarkers in combination, specifically IL-8, MIP-1β, and GM-CSF.

That IL-8 was identified as the most predictive biomarker in both systems is not surprising; its utility and precedence as a predictive biomarker of sensitization have already been established. MIP-1β is a chemoattractant for a variety of inflammatory cells including T-cells, natural killer cells, monocytes, and macrophages. It has also been shown to be involved in T-cell trafficking into lymph nodes and has previously been studied as a biomarker of sensitization in THP-1 cells with moderate success[36]. Keratinocytes are known to produce GM-CSF during the pro-inflammatory cascade triggered by the early stages of DC activation[2]. This may explain why GM-CSF is more predictive for the tri-culture system

than the MUTZ-LC mono-culture. Overall, the classification models developed in the SVM demonstrate that multiple metrics combined offer increased accuracy over single metrics. It was also shown that the tri-culture system outperforms MUTZ alone in prediction of skin sensitizers, pre- and pro-haptens in particular. It is interesting to note that the most accurate molecular signature for pre- and pro-hapten sensitizer prediction differs from that of overall prediction. This may correlate with an increased role for keratinocytes in the sensitization process by pre- and pro-haptens. While DC-like cells have been shown to secrete MIP-1β, keratinocytes have been shown to express RANTES[15]. On the other hand, the biomarker ranking for MUTZ-LC prediction of pre- and pro-haptens did not change and neither did the accuracy, reflective of a more static mono-culture environment. We further demonstrated that the biomarker panel identified for the tri-culture system achieves similar accuracy (91%), sensitivity (92.7%), and specificity (89.8%) of prediction of skin sensitizers compared to the full-thickness skin model (accuracy, sensitivity, and specificity all 92%)[22] (Supplementary Table 1). This suggests that the tri-culture system could offer a suitable and more cost-effective alternative to full-thickness skin models without sacrificing accuracy.

Classification trees (c-trees) are a model of decision tree learning. The conditional interference approach employed by the algorithm is used to avoid the issue of biased predictor selection. The resulting trees identify a node(s) that specifies a statistical threshold. The first node/threshold identified in our classification trees is used to distinguish between all sensitizer potencies and non-sensitizers. Using these thresholds, it was easy to identify the misclassified chemicals. The MUTZ-LCs alone identified a number of false negatives while the tri-culture system had no false negatives. The chemicals identified as false positives by the c-tree for the tri-culture system warrant notice; in particular, all three concentrations of vanillin were categorized as sensitizers. Vanillin has been shown to be weakly sensitizing in humans and guinea pigs and was identified as a "false" positive in U-SENS™, a THP-1 study using ROS to predict sensitization, and an *in silico* combined test strategy using h-CLAT, DPRA, and DEREK[37–39]. It stands to reason that vanillin could indeed be a weakly sensitizing agent and not a non-sensitizer, in which case, there was no error in classification.

The second node in the c-tree produced a threshold to distinguish between weak/moderate (W/M) and strong/extreme (S/E) sensitizers. These potency groups were chosen based on the current GHS subcategories 1A and 1B (1A is S/E, 1B is W/M)[40]. The tri-culture system correctly classified all W/M sensitizers with overall accuracy of 83%. This is in direct contrast to animal models, which have been shown to produce false negatives for weak sensitizers based on the classification decision threshold. The GPMT in particular had a 30%-sensitized cutoff below which a chemical was classified as a non-sensitizer when it was in fact weakly sensitizing[41]. Overall, the c-tree method demonstrates promise for potency prediction using cytokine SI, but requires further development and validation with a larger panel of chemicals. An expanded chemical panel may further allow for sub-division of the weak/moderate and strong/extreme potency categories. It is possible that combining prediction with MIP-1β and GM-CSF will improve potency prediction.

In conclusion, a tri-culture system was developed that utilizes keratinocytes, fibroblasts, and MUTZ-LCs as a potential alternative to animal testing for identification of all sensitizer

classes. The utility of IL-8 as a starting biomarker of sensitization was demonstrated. Furthermore, a multi-metric predictive signature for this approach was established using a support vector machine to rank the margin distances from a panel of 27 secreted cellular metrics. This assay offers the metabolic capabilities of skin to convert pre- and pro-hapten sensitizers into electrophilic products and predict their sensitization potential and potency. Future studies will focus on optimizing incubation time points and viability assessment, expanding the chemical panel further, and identifying other predictive metrics beyond the secretome. With optimization and validation, we can introduce a high-accuracy, cost-effective, high-throughput, multi-metric assay for screening skin sensitizers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

1. Thyssen JP et al. The epidemiology of contact allergy in the general population — Prevalence and main findings. Contact Dermatitis 57(5), 287–299 (2007). [PubMed: 17937743]

2. Enk AH Allergic contact dermatitis: Understanding the immune response and potential for targeted therapy using cytokines. Mol. Med. Today 3(10), 423–428 (1997). [PubMed: 9358468]

3. Divkovic M et al. Hapten-protein binding: From theory to practical application in the in vitro prediction of skin sensitization. Contact Dermatitis 53(4), 189–200 (2005). [PubMed: 16191014]

4. Peiser M et al. Allergic contact dermatitis: Epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. Current knowledge assembled at an international workshop at BfR, Germany. Cell Mol. Life Sci 69(5), 763–781 (2012). [PubMed: 21997384]

5. Karlberg AT et al. Allergic contact dermatitis — Formation, structural requirements, and reactivity of skin sensitizers. Chem. Res. Toxicol 21(1), 53–69 (2008). [PubMed: 18052130]

6. Anderson SE et al. The LLNA: A brief review of recent advances and limitations. J. Allergy (Cairo) 2011, 424203 (2011). [PubMed: 21747867]

7. Chipinda I et al. Use of the human monocytic leukemia THP-1 cell line and co-incubation with microsomes to identify and differentiate hapten and prohapten sensitizers. Toxicology 280(3), 135–143 (2011). [PubMed: 21163322]

8. Hennen J et al. Cross talk between keratinocytes and dendritic cells: Impact on the prediction of sensitization. Toxicol. Sci 123(2), 501–510 (2011). [PubMed: 21742781]

9. Bergstrom MA et al. A skin-like cytochrome P450 cocktail activates prohaptens to contact allergenic metabolites. J. Invest. Dermatol 127(5), 1145–1153 (2007). [PubMed: 17124504]

10. Gotz C et al. Xenobiotic metabolism capacities of human skin in comparison with a 3D epidermis model and keratinocyte-based cell culture as in vitro alternatives for chemical testing: Activating enzymes (Phase I). Exp. Dermatol 21(5), 358–363 (2012). [PubMed: 22509833]

11. Rolsted K et al. Evaluation of cytochrome P450 activity in vitro, using dermal and hepatic microsomes from four species and two keratinocyte cell lines in culture. Arch. Dermatol. Res 300(1), 11–18 (2008). [PubMed: 17992533]

12. Hagvall L et al. Fragrance compound geraniol forms contact allergens on air exposure. Identification and quantification of oxidation products and effect on skin sensitization. Chem. Res. Toxicol 20(5), 807–814 (2007). [PubMed: 17428070]

13. Hagvall L et al. Cytochrome P450-mediated activation of the fragrance compound geraniol forms potent contact allergens. Toxicol. Appl. Pharmacol 233(2), 308–313 (2008). [PubMed: 18824010]

14. Kalmes M & Blomeke B Impact of eugenol and isoeugenol on AhR translocation, target gene expression, and proliferation in human HaCaT keratinocytes. J. Toxicol. Environ. Health A 75(8-10), 478–491 (2012). [PubMed: 22686307]

15. Corsini E & Galli CL Epidermal cytokines in experimental contact dermatitis. Toxicology 142(3), 203–211 (2000). [PubMed: 10667891]

16. Nishibu A et al. Roles for IL-1 and TNFalpha in dynamic behavioral responses of Langerhans cells to topical hapten application. J. Dermatol. Sci 45(1), 23–30 (2007). [PubMed: 17123788]

17. Cumberbatch M et al. Tumour necrosis factor-alpha induces Langerhans cell migration in humans. Br. J. Dermatol 141(2), 192–200 (2009).

18. Mascia F et al. EGFR regulates the expression of keratinocyte-derived granulocyte/macrophage colony-stimulating factor in vitro and in vivo. J. Invest. Dermatol 130(3), 682–693 (2010). [PubMed: 19890352]

19. Ouwehand K et al. CXCL12 is essential for migration of activated Langerhans cells from epidermis to dermis. Eur. J. Immunol 38(11), 3050–3059 (2008). [PubMed: 18924211]

20. Saalbach A et al. Dermal fibroblasts induce maturation of dendritic cells. J. Immunol 178(8), 4966–4974 (2007). [PubMed: 17404278]

21. Luu-The V et al. Expression profiles of phases 1 and 2 metabolizing enzymes in human skin and the reconstructed skin models Episkin and full thickness model from Episkin. J. Steroid Biochem. Mol. Biol 116(3-5), 178–186 (2009). [PubMed: 19482084]

22. Lee S et al. Predicting full thickness skin sensitization using a support vector machine. Toxicol. In Vitro 28(8), 1413–1423 (2014). [PubMed: 25025180]

23. Asturiol D et al. Consensus of classification trees for skin sensitisation hazard prediction. Toxicol In Vitro 36, 197–209 (2016). [PubMed: 27458072]

24. Vandebriel RJ & van Loveren H Non-animal sensitization testing: State-of-the-art. Crit. Rev. Toxicol 40(5), 389–404 (2010). [PubMed: 20180632]

25. Ouwehand K et al. Comparison of a novel CXCL12/CCL5 dependent migration assay with CXCL8 secretion and CD86 expression for distinguishing sensitizers from non-sensitizers using MUTZ-3 Langerhans cells. Toxicol. In Vitro 24(2), 578–585 (2010). [PubMed: 19878716]

26. Nelissen I et al. MUTZ-3-derived dendritic cells as an in vitro alternative model to CD34+ progenitor-derived dendritic cells for testing of chemical sensitizers. Toxicol. In Vitro 23(8), 1477–1481 (2009). [PubMed: 19732821]

27. Kimura Y et al. Optimization of the IL-8 Luc assay as an in vitro test for skin sensitization. Toxicol. In Vitro 29(7), 1816–1830 (2015). [PubMed: 26187477]

28. Larsen CG et al. Production of interleukin-8 by human dermal fibroblasts and keratinocytes in response to interleukin-1 or tumour necrosis factor. Immunology 68(1), 31–36 (1989). [PubMed: 2478449]

29. Taub DD et al. T lymphocyte recruitment by interleukin-8 (IL-8). IL-8-induced degranulation of neutrophils releases potent chemoattractants for human T lymphocytes both in vitro and in vivo. J. Clin. Invest 97(8), 1931–1941 (1996). [PubMed: 8621778]

30. Larsen CG et al. The delayed-type hypersensitivity reaction is dependent on IL-8. Inhibition of a tuberculin skin reaction by an anti-IL-8 monoclonal antibody. J. Immunol 155(4), 2151–2157 (1995). [PubMed: 7636263]

31. Takahashi T et al. An in vitro test to screen skin sensitizers using a stable THP-1-derived IL-8 reporter cell line, THP-G8. Toxicol. Sci 124(2), 359–369 (2011). [PubMed: 21920952]

32. Toebak MJ et al. CXCL8 secretion by dendritic cells predicts contact allergens from irritants. Toxicol. In Vitro 20(1), 117–124 (2006). [PubMed: 16099135]

33. Bauch C et al. Putting the parts together: Combining in vitro methods to test for skin sensitizing potentials. Regul. Toxicol. Pharmacol 63(3), 489–504 (2012). [PubMed: 22659254]

34. Almeida A et al. Insights on in vitro models for safety and toxicity assessment of cosmetic ingredients. Int. J. Pharm 519(1-2), 178–185 (2017). [PubMed: 28104405]

35. Johansson H et al. A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. BMC Genomics 12, 399 (2011). [PubMed: 21824406]

36. Hirota M & Moro O MIP-1beta, a novel biomarker for in vitro sensitization test using human monocytic cell line. Toxicol. In Vitro 20(5), 736–742 (2006). [PubMed: 16314067]

37. Piroird C et al. The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. Toxicol In Vitro 29(5), 901–916 (2015). [PubMed: 25820135]

38. Saito K et al. Development of an *in vitro* skin sensitization test based on ROS production in THP-1 cells. Toxicol. In Vitro 27(2), 857–863 (2013). [PubMed: 23287046]

39. Takenouchi O et al. Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. J. Appl. Toxicol 35(11), 1318–1332 (2015). [PubMed: 25820183]

40. Committee on the Design and Evaluation of Safer Chemical Substitutions: A Frame-work to Inform Government and Industry Decision; Board on Chemical Sciences and Technology; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Research Council. A Framework to Guide Selection of Chemical Alternatives. Washington (DC): National Academies Press (US); 2014 10 29 Appendix D, Overview of the GHS Classification Scheme in Hazard Classification. Available from: https://www.ncbi.nlm.nih.gov/books/NBK253967/

41. Basketter DA & Kimber I Skin sensitization, false positives and false negatives: Experience with guinea pig assays. J. Appl. Toxicol 30(5), 381–386 (2010). [PubMed: 20583317]
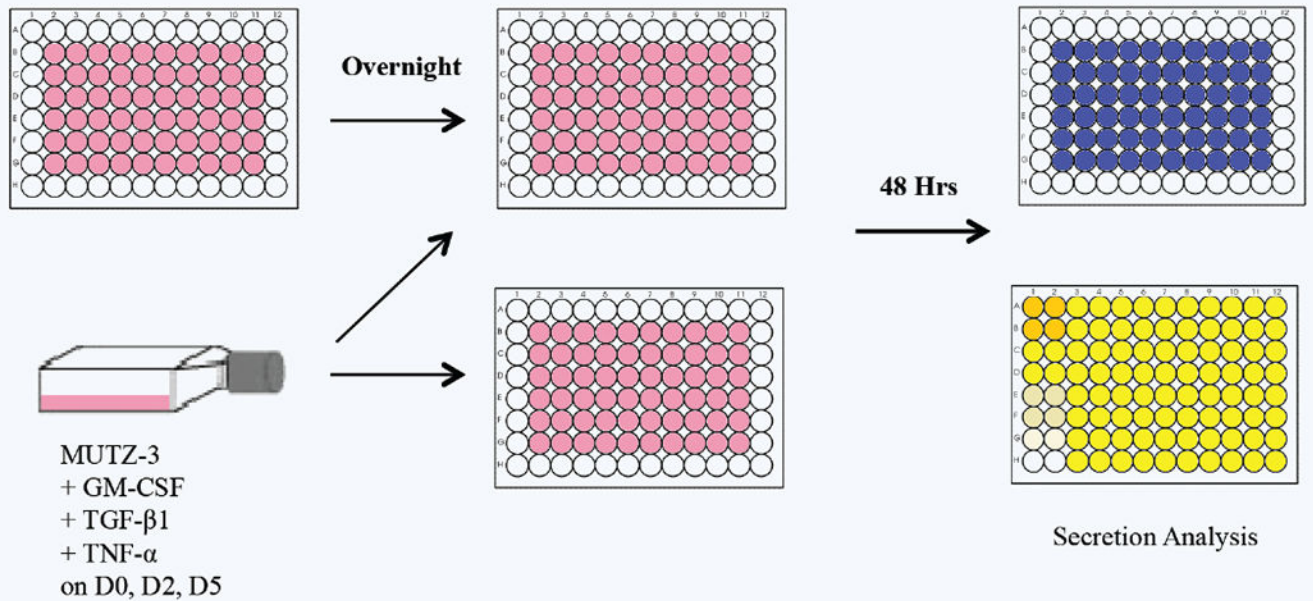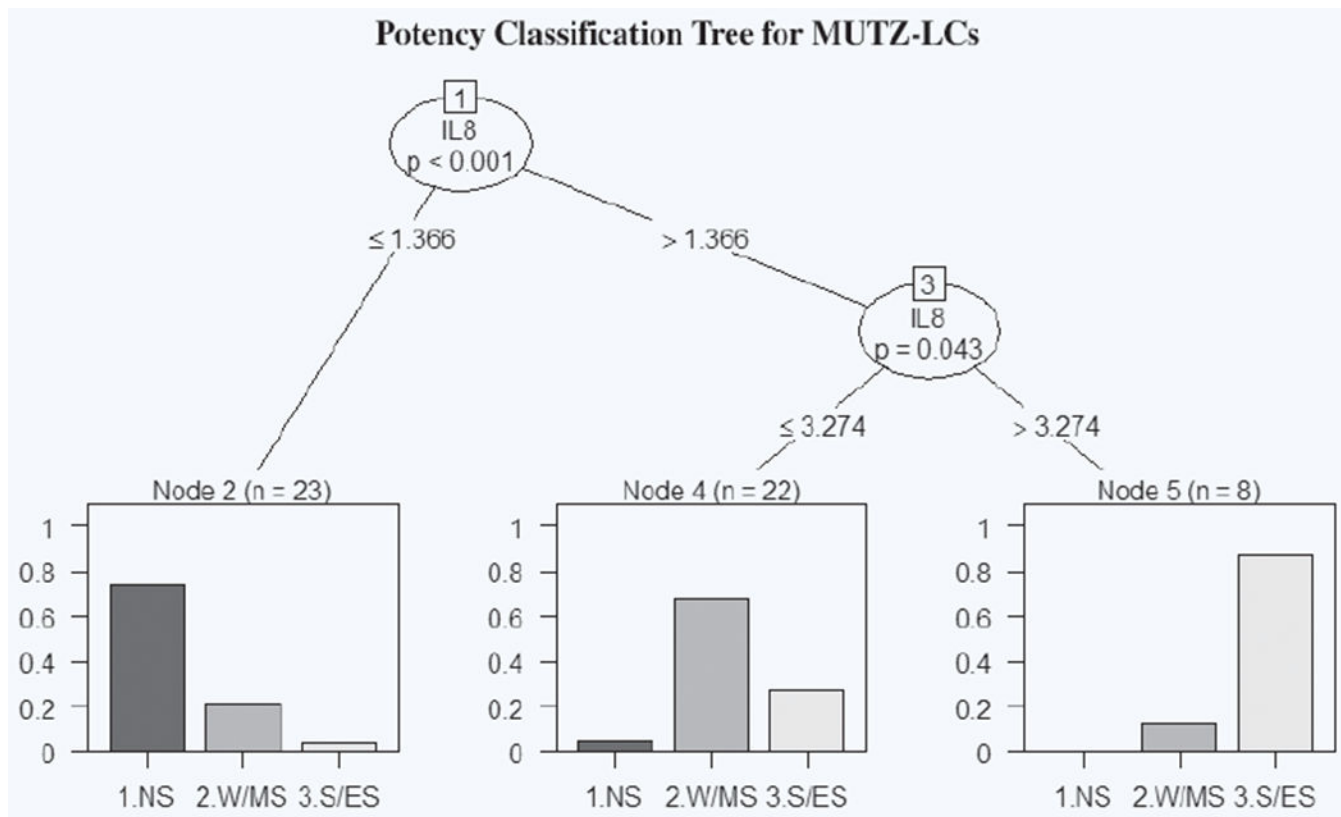
Figure 1.
Experimental schematic for tri-culture and chemical treatment of MUTZ-LCs, HaCaT KCs, and dermal FBs and MUTZ-LCs in mono-culture from days 0-9. MUTZ-3 cells were cultured in complete media and growth factors on D0, D2, and D5. On D6 KCs and FBs were plated to adhere overnight. On D7 MUTZ-LCs were plated in tri-culture and mono-culture and dosed with the chemicals listed in Table 1; 48 hours after dosing, supernatant was collected from each culture system for secretome analysis and parallel plate was used for Alamar Blue viability analysis.
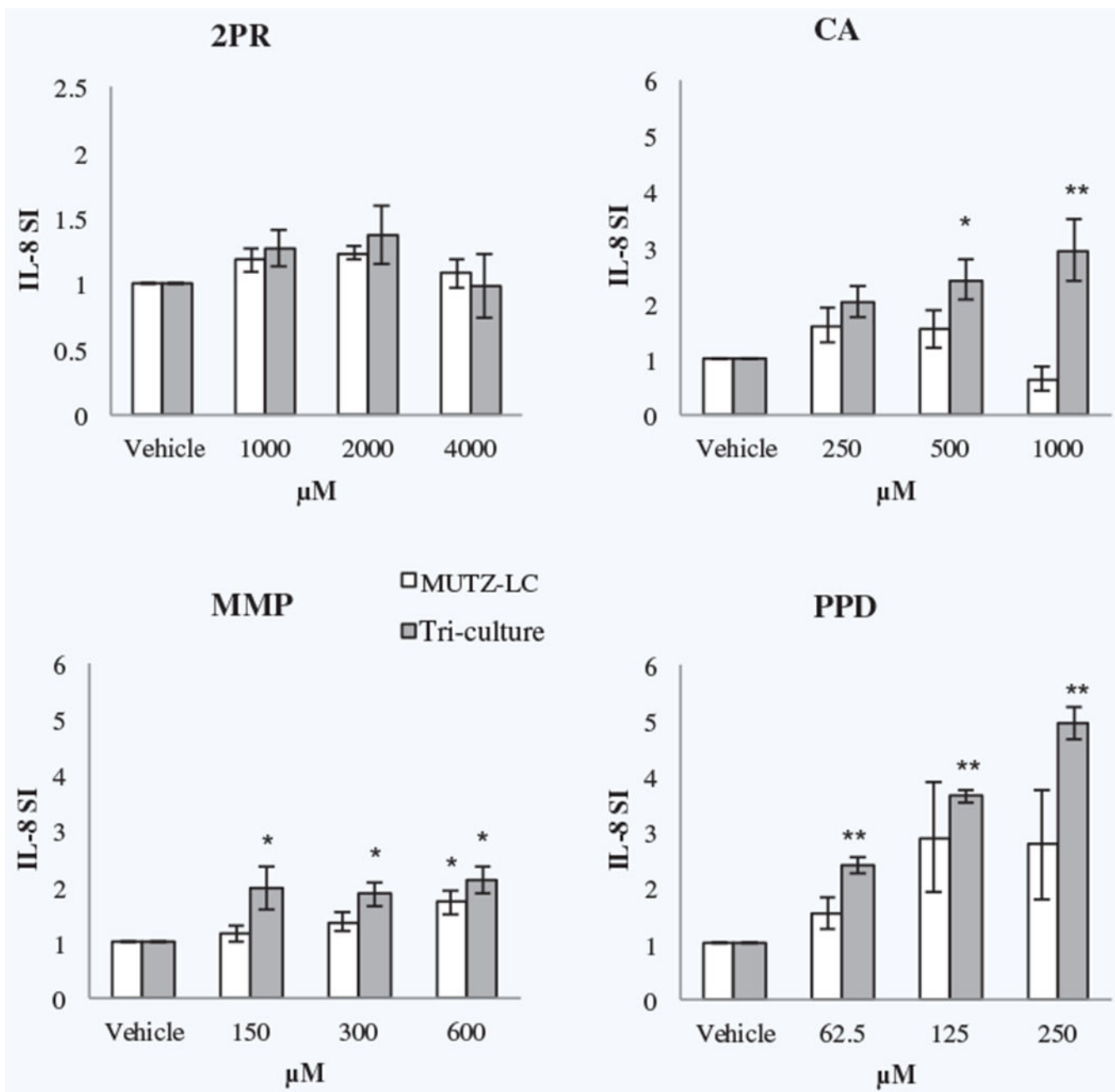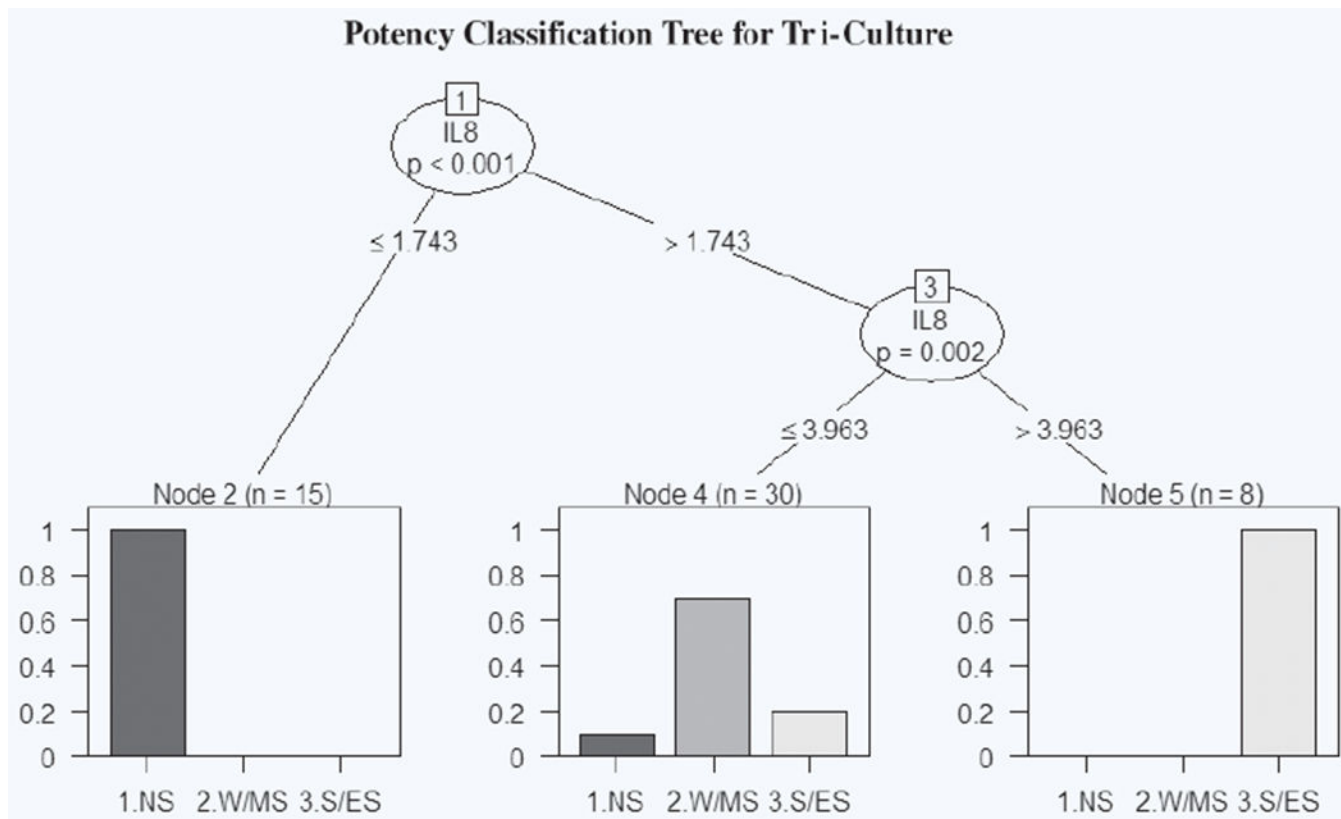
Figure 2.
ELISA analysis of IL-8 secretion by the tri-culture system and MUTZ-LCs alone in response to non-sensitizer isopropanol and pre-/pro-hapten sensitizers cinnamic alcohol (weak), 2-methoxy-4-methylphenol (moderate), and *p*-phenylenediamine (strong). * indicates p 0.05 and ** indicates p 0.005 by ANOVA, Fisher's LSD post-hoc analysis for $n = 4$ independent replicates.

Figure 3.
IL-8 potency classification tree generated in R for MUTZ-LCs. The first node splits at SI = 1.366 and classifies the chemical based on SI value as non-sensitizer (SI ≤ 1.366) or sensitizer (SI > 1.366). The second node splits at SI = 3.274 and classifies sensitizers based on SI value as W/M (SI ≤ 3.274) or S/E (SI > 3.274). Each bar graph represents the percentage breakdown of chemicals classified in each category by their known classification. "*n*" represents chemical concentrations grouped in each leaf.

Figure 4.

IL-8 potency classification tree generated in R for tri-culture. The first node splits at SI = 1.743 and classifies the chemical based on SI value as non-sensitizer (SI   1.743) or sensitizer (SI > 1.743). The second node splits at SI = 3.963 and classifies sensitizers based on SI value as W/M (SI   3.963) or S/E (SI > 3.963). Each bar graph represents the percentage breakdown of chemicals classified in each category by their known classification. "*n*" represents chemical concentrations grouped in each leaf.

Table 1

Chemical panel evaluated.

Chemical	Abbreviation	Class	Potency	Concentrations (μM)
Sensitizers				
2-4-dinitrochlorobenzene	DNCB	Hapten	Extreme	6.25, 12.5
p-benzoquinone	pBQ	Pre-/Pro-Hapten	Extreme	10, 25, 50
2-aminophenol	2AP	Pre-/Pro-Hapten	Strong	100, 200, 400
Hydroquinone	HQ	Pre-/Pro-Hapten	Strong	25, 50, 100
p-phenylenediamine	PPD	Pre-/Pro-Hapten	Strong	62.5, 125, 250
Cinnamaldehyde	CLD	Hapten	Moderate	62.5, 125, 250
Isoeugenol	IE	Pre-/Pro-Hapten	Moderate	250, 500, 1000
2-methoxy-4-methylphenol	MMP	Pre-/Pro-Hapten	Moderate	150, 300, 600
Resorcinol	RC	Pre-/Pro-Hapten	Moderate	500, 1000, 2000
Cinnamic Alcohol	CA	Pre-/Pro-Hapten	Weak	250, 500, 1000
Eugenol	EU	Pre-/Pro-Hapten	Weak	250, 500, 1000
Geraniol	GER	Pre-/Pro-Hapten	Weak	250, 500, 1000
Non-sensitizers				
Dimethylsulfoxide	DMSO	Vehicle control	—	0.10%
Isopropanol	2PR	Non-sensitizer	—	1000, 2000, 4000
Lactic Acid	LA	Irritant	—	250, 500, 1000
Salicylic Acid	SA	Irritant	—	500, 1000, 2000
Sodium Dodecyl Sulfate	SDS	Irritant	—	125, 250
Vanillin	VL	Irritant	—	125, 250
Xylene	XYL	Non-sensitizer	—	750, 1500, 3000

Table 2

SVM prediction accuracy scores.

MUTZ-LCs alone		Tri-culture	
Metric	*Accuracy*	*Metric*	*Accuracy*
Top 12	**87.2%**	Top 12	86.7%
Top 4 (IL-8, MIP-1β, IL-9, IL-17)	85%	Top 4 (IL-8, MIP-1β, GM-CSF, RANTES)	90%
Top 3 (IL-8, MIP-1β, IL-9)	83%	Top 3 (IL-8, MIP-1β, GM-CSF)	**91.1%**
Top 2 (IL-8, MIP-1β)	86.2%	Top 2 (IL-8, GM-CSF)	87.8%
IL-8	84%	IL-8	86.7%

SVM accuracy for both *in vitro* systems using feature selection. The highest achievable accuracy for each system is noted in bold print; for both systems, the highest accuracy was achieved in combination rather than by the top-ranking biomarker, IL-8, alone.

Table 3

C-tree prediction accuracy scores.

	MUTZ-LC	
	Sensitivity	*Specificity*
Non-sensitizers	73.9%	96.7%
Weak/moderate sensitizers	68.2%	80.6%
Strong/extreme sensitizers	87.5%	84.4%
Overall accuracy: 73.6%		

	Tri-culture	
	Sensitivity	*Specificity*
Non-sensitizers	100%	92.1%
Weak/moderate sensitizers	70.0%	100%
Strong/extreme sensitizers	100%	86.7%
Overall accuracy: 83.0%		

C-tree accuracy, sensitivity, and specificity of prediction of sensitizer potency for both *in vitro* systems.