AN IN VITRO METHOD FOR MULTI-METRIC ANALYSIS OF SKIN SENSITIZERS

by

TALIA M. GREENSTEIN

A thesis submitted to the

Graduate School – New Brunswick

And

The Graduate School of Biomedical Sciences

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Biomedical Engineering

Written under the direction of

Martin L. Yarmush

And approved by

_____

_____

_____

New Brunswick, New Jersey

May 2017

ABSTRACT OF THE THESIS

AN IN VITRO METHOD FOR MULTI-METRIC ANALYSIS OF SKIN SENSITIZERS

by TALIA M. GREENSTEIN

Thesis Direction:
Martin L. Yarmush, MD, PhD

Allergic contact dermatitis is an inflammatory skin disease that impacts 15-20% of the general population. The prevalence of potentially sensitizing agents necessitates screening methods for chemical risk assessment. Variability, cost, ethical concerns, and the recent ban of animal testing in the European Union introduce a pressing need to develop alternative methods to animal testing to screen for the sensitization potential of chemicals. Many current screening approaches are limited in their ability to predict pre- and pro-hapten sensitizers, which require abiotic or metabolic conversion prior to inducing sensitization, and sensitizer potency. We previously developed an *in vitro* co-culture system of MUTZ-3-derived Langerhans cells, HaCaT keratinocytes, and primary dermal fibroblasts to mimic the *in vivo* cellular and metabolic environments of skin sensitization. We expanded the chemical test panel to include a variety of non-sensitizers, haptens, pre-haptens, and pro-haptens of all potencies and compared the co-culture system's performance to MUTZ-3 Langerhans cells alone by measuring CXCL8 secretion. The secretome of both cultures were also evaluated for 27 cytokines, chemokines, and growth factors. A support vector machine was used to identify the most predictive signature of sensitization for each culture system and classification trees were used to identify statistical thresholds to predict sensitizer potency by CXCL8 secretion. The support vector machine computed prediction accuracy of 87% for the MUTZ-3 mono-culture system using the top 12 ranked biomarkers while predicting accuracy of 91% for the co-culture system using the top 3 (IL-8, MIP-1β, and GM-CSF). The classification trees demonstrated 83% accuracy for potency

prediction by the co-culture system and 73% accuracy for MUTZ-3 cells alone. Overall, the presence of keratinocytes and fibroblasts enhanced the system's ability to detect pre- and pro-haptens, resulting in higher accuracy scores for the co-culture system. To further reduce the use of animal derivatives, other predictive metrics are being explored. Further studies using the cells of our co-culture demonstrate that the Langerhans cells are requisite for CXCL8 secretion, and therefore RNA expression in sensitized Langerhans cells from co-culture will be assessed to identify additional predictive metrics. Mitochondrial markers of cellular stress due to sensitization are also being pursued in this system. As such, this co-culture *in vitro* assay presents a promising alternative method to animal testing for screening and classifying potential skin sensitizers and offers the possibility of multi-metric analysis for enhanced screening capabilities and mechanistic studies of allergic contact dermatitis.

DEDICATION

This thesis is dedicated to three incredible women

to Emily Blady, who led me to my path

to Dr. Rene Schloss, who met me at the beginning of it and walks with me still

and

to Dr. Serom Lee, who taught me all the steps along the way.

ACKNOWLEDGEMENTS

I would like to thank the members of my committee: Dr. Martin Yarmush, Dr. Rene Schloss, and Dr. Nada Boustany. They have each contributed to the development and completion of this thesis in unique ways. I would like to thank my advisor, Dr. Yarmush, for affording me a place in this lab, where I have had the opportunity to grow and flourish as a scientist. His guidance over the years has been instrumental in my decision to pursue higher education and research. As I prepare to follow in her footsteps, I am especially grateful to Dr. Rene Schloss for all the time together in meetings, the chocolate, and for introducing me to the world of immunology. Her words of wisdom, guidance, and support helped shape the person I am today and the career path I have chosen. I would also like to thank Dr. Boustany for sharing her expertise in mitochondrial imaging and her guidance through this project.

I am so grateful to be part of a lab that is like family to me. I would like to extend my thanks to Dr. Serom Lee, Ileana Marrero-Berrios, Mollie Davis, Paulina Krzyszczyk, Isabel Perez, Lingting Shi, Dr. Andrea Gray, Hwan June Kang, Zachary Fritz, Dr. Tim Maguire, and Dr. Suneel Kumar. I want to thank Dr. Serom Lee in particular for demonstrating by example what a great mentor is. She has played a critical role in my growth as a scientist and as a person, as I found myself looking up to her like a big sister. Dr. Tim Maguire has played a critical role in developing and troubleshooting the computational methods used in this project. I would also like to thank Lingting Shi for all the work she has put into this project.

I am grateful to Emily Blady for making the “shidduch” between Rene and me; that is something I can never repay.

I would not have reached this milestone without the love and support from my family. I am grateful to my parents for instilling in me a passion for science and thirst for knowledge. Their unwavering support helped me reach this point. They taught me to persist and to reach for the stars, and I hope to make them proud as I continue on this journey. I would like to thank my mother- and father-in-law for their support and encouragement. Last, but not least, I would like

to thank my husband, Gavi Dov, for inspiring me to be a better person every day through his patience, love, and kindness, and for teaching me to think logically about everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 4: THESIS CONCLUSIONS AND FUTURE WORK

ABBREVIATIONS

2AP – 2-aminophenol

2PR – Isopropanol

ACD – Allergic Contact Dermatitis

AOP – Adverse Outcome Pathway

c-tree – Classification Tree

CA – Cinnamic Alcohol

CLD – Cinnamaldehyde

DC – Dendritic cell

DPRA – Direct Peptide Reactivity Assay

DMSO – Dimethylsulfoxide

DNCB – 2-4-dinitrochlorobenzene

ECVAM – European Center for the Validation of Alternative Methods

ELISA – Enyzme-linked Immunosorbent Assay

EU – Eugenol

FB – Fibroblast

GER – Geraniol

GPMT – Guinea Pig Maximization Test

HQ – Hydroquinone

ICCVAM – Interagency Coordinating Committee on the Validation of Alternative Methods

ICD – Irritant Contact Dermatitis

IE – Isoeugenol

IL – Interleukin

KC – Keratinocyte

LA – Lactic Acid

LC – Langerhans cell

LLNA – Local Lymph Node Assay

MMP – 2-methoxy-4-methylphenol

MUTZ-LCs – MUTZ-3-derived Langheran-like cells

NS – Non-sensitizer

OCD – Occupational Contact Dermatitis

OECD – Organization for Economic Cooperation and Development

pBQ – p-benzoquinone

PPD – p-phenylenediamine

QSAR – Quantitative Structure Activity Relationship

RC – Resorcinol

S – Sensitizer

SA – Salicylic Acid

SDS – Sodium Dodecyl Sulfate

S/E – Strong/Extreme

SI – Stimulation Index

SVM – Support Vector Machine

VL – Vanillin

W/M – Weak/Moderate

XYL – Xylene

CHAPTER 1: INTRODUCTION

1.1 ALLERGIC CONTACT DERMATITIS

Contact dermatitis (CD) is generally defined as either an acute or chronic inflammatory reaction to irritants or sensitizers. CD accounts for 80-90% of all occupational skin diseases. Allergic Contact Dermatitis (ACD), a subtype of contact dermatitis, is a delayed hypersensitivity response that prompts a T-cell mediated response. It is currently one of the most prevalent dermatoses, accounting for 60% of all occupational contact dermatitis, and costing at least $1 billion annually in the United States in associated medical services, treatments, and loss of productivity. Approximately 20% of the general population will be impacted by contact allergy to at least one allergen. There is increased prevalence of contact allergy among women (11.8-35.4%) compared to men (6-18%), likely due to differences in exposure and intrinsic sex differences in skin susceptibility. Nickel, thimerosal, chromium, p-phenylenediamine (PPD), and fragrance mixes are among the most common allergens[1-3].

ACD is clinically characterized by pruritus and erythematous reactions, generally located where the allergen was applied. The sensitized area may become vesiculous or exudative and is typically very itchy. The first step of ACD is the sensitization phase, in which a sensitizing molecule penetrates the skin. The allergenic hapten-immunogen complex binds to epidermal Langerhans cells (LCs), thus triggering primary cytokine production (i.e. IL-1β) to induce a pro-inflammatory cytokine and chemokine cascade response in neighboring keratinocytes (KCs). Among these cytokines and inflammatory mediators are tumor necrosis factor α (TNF-α), IL-1α, IL-6, IL-10, granulocyte-macrophage colony-stimulating factor (GM-CSF), reactive oxygen species (ROS), IL-18, and IL-8. The cytokine cascade induces maturation of LCs while the chemokines attract other immune cells to the area. LCs are a dendritic cell subset that function as potent antigen-presenting cells in this reaction. Mature LCs up-regulate co-stimulatory factors and MHC I and II molecules to present to specific T-cells. Hapten-bearing LCs migrate to the regional lymph nodes, where antigen-specific T-cells are primed and begin clonal expansion of effector and memory T-

cells in response. These T cells migrate to the blood and the area of sensitized skin and trigger a second wave of leukocyte infiltration, including neutrophils, T-cells, and inflammatory monocytes. Resolution of ACD involves clearing the haptens from the skin and activation of regulatory T cells (Tregs). The sensitization step in the first exposure to a contact allergen lasts 10-15 days in humans, and often has no clinical consequences. Re-exposure to the same hapten in an individual leads to the appearance of ACD, generally within 24-72 hours[3-6]. Repeated contact with a sensitizing agent may result in the development of a more chronic and less inflammatory form of ACD. There is no cure for ACD; treatment involves anti-inflammatory corticosteroids to relieve the symptoms. Diagnosis of ACD is generally confirmed by patch testing[7].

It should be noted that occupational skin diseases include a number of distinct dermatoses, including allergic contact dermatitis (ACD), atopic dermatitis, and irritant contact dermatitis (ICD). ICD, like ACD, is caused by an external agent. However, ICD differs from ACD in that it induces an innate, or non-allergic, inflammatory response to the irritant molecule. ICD does not induce a T-cell mediated response, and thus repeated exposure to the irritant agent does not pose risk of systemic response[1].

1.2 SENSITIZERS

Landsteiner and Jacobs postulated that small organic molecules complex with skin proteins to form the active immunogen in skin sensitization in 1935[8]. These organic molecules are too small (typically <500 Da) to be recognized by the classical immunological mechanisms, and therefore must be complexed with skin proteins and produce covalent adducts in order to elicit an immune response[9]. The most common functional reaction groups in sensitizing agents are Michael acceptors, acylating agents, Schiff base formers, S_NAr electrophiles, and S_N1/S_N2 electrophiles[7].

There are three classes of contact allergens, which are grouped according to the mechanistic pathways through which they form macro-molecular immunogens that may initiate an allergic response (**Figure 1.1**). The first class, haptens, are small, generally electrophilic sensitizers that

bind readily to covalently modify nucleophilic skin proteins to form an immunogenic macromolecule. Some non-electrophilic haptens directly haptenate proteins via disulfide formation. Metal allergens represent non-classical haptens in that they form metal-protein complexes that are not sufficiently strong to survive antigen processing like classical haptens, and appear to bypass intracellular antigen processing steps[10]. The remaining two classes of contact allergens, pre- and pro-haptens, are inherently non-reactive until their structure is modified. Pre-haptens undergo external, abiotic mechanisms of activation such as auto-oxidation or photoactivation to form reactive molecules that may bind to skin proteins and form a sensitizing entity[11]. Pro-haptens require metabolic activation in the skin to form reactive electrophilic intermediaries that can bind skin proteins. Multiple enzymes are involved in the activation of pro-haptens, the most prominent of which is the cytochrome P450 family[12]. Normal human epidermal keratinocytes and dermal fibroblasts have been shown to express CYP1A1, 1B1, 2B6, 2E1, and 3A5. Other enzymes involved in metabolic conversion of pro-haptens include alcohol dehydrogenases, aldehyde dehydrogenases, monoamine oxidases, flavin-containing monooxygenases, hydrolytic enzymes, acyltransferases, glutathione S-transferases, uridine 5'-diphospho-glucuronosyltransferases and sulphotransferases[10,13]. Several chemicals may act as both pre- and pro-haptens; that is, one chemical may undergo autoxidation or enzymatic activation and form the same hapten. Geraniol and alpha-terpinene are examples of such chemicals whose secondary oxidation products were found to be identical to those formed via the metabolic pathway[14].

In 2003, the United Nations adopted the Globally Harmonized System of Classification and Labeling of Chemicals (GHS). The GHS includes criteria for the classification of health and environmental hazards and specifies what information should be included on labels of hazardous chemicals and their associated safety data sheets. Skin sensitizers are classified as GHS category 1, and can be further classified into subcategories 1A or 1B based on potency assessment data, if available. Chemicals in subcategory 1A show high frequency of sensitization occurrence in

humans and/or a high potency in animals and are presumed to potentially produce significant sensitization in humans. Chemicals in subcategory 1B show low to moderate frequency of sensitization occurrence in humans and/or low to moderate potency in animals. The evidence used to classify sensitizers includes positive data from patch testing and/or animal studies, and documented episodes of ACD, generally obtained from dermatology clinics[15]. Potency assessment is important, as the resulting data contributes to hazard classification and can lead to improvements in risk management. Currently, there is no streamlined method of skin sensitizer potency assessment.

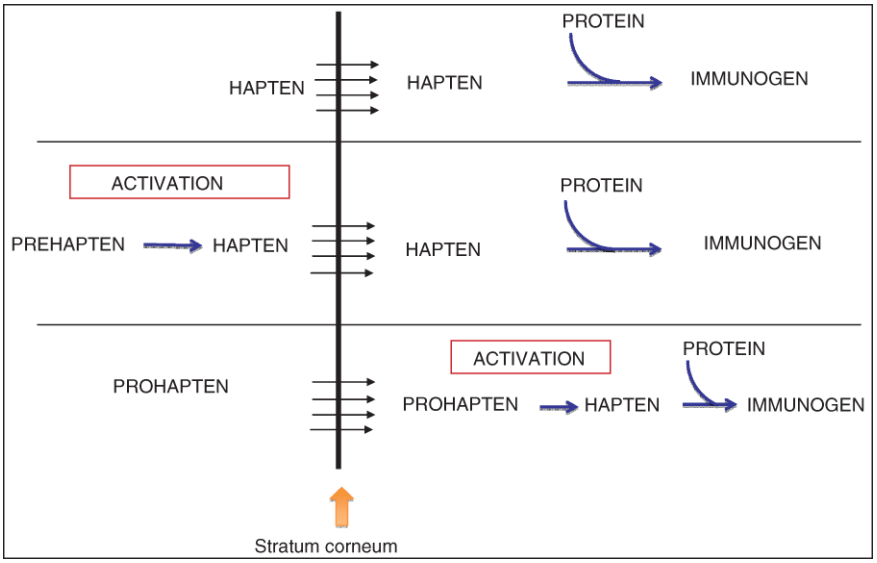

Figure 1.1. Three classes of contact allergens and their mechanisms of forming active immunogens. Haptens pass through the stratum corneum and readily bind skin proteins to form an active immunogen. Pre-haptens are activated outside the skin to form haptens that can bind skin proteins. Pro-haptens are metabolically activated in the skin to form haptens that can bind skin proteins. Reprinted with permission from Karlberg *et al*, Contact Dermatitis[11] © 2013 John Wiley & Sons A/S.

As of 2008, there were 5,288 individual substances registered in the New Chemicals Database (managed by the former European Chemicals Bureau), of which 3,792 were reported to have been tested for posed hazard of skin sensitization. 1,047 of those tested were determined to pose risk of skin sensitization[16]. New chemicals are continuously submitted to the European Chemicals Agency for registration. As such, a streamlined method of screening new and old chemicals for their potential risk to the general population is crucial to public and occupational safety.

1.3 CURRENT SCREENING APPROACHES

Many of the chemicals used in everyday cosmetics may be sensitizing agents and therefore must be tested to determine if they are safe (and if so, at what concentration) for distribution among the general population. Indeed, two fragrance mixes have been used as screening agents for contact allergy in baseline patch test series. Fragrance mix I consists of aromatic sensitizers amyl cinnamal, cinnamal, cinnamyl alcohol, eugenol, geraniol, hydroxycitronellal, isoeugenol, and oakmoss absolute. Fragrance mix II consists of citral, citronellol, coumarin, farnesol, hexyl cinnamal, and hydroxyisohexyl 3-cyclohexene carboxaldehyde[17]. Many of these chemicals are currently known to induce high incidence of ACD.

1.3.1 *IN VIVO* (ANIMAL) TESTING

Early Screening Methods

The advances of synthetic chemistry and increases in medical, chemical, and technical knowledge in the early 1900's introduced an unprecedented attack on human skin that demanded attention. Landsteiner and Jacobs established the white male guinea pig as a suitable test model for skin sensitization studies in 1935[8]. The Schwartz-Peck Test was the first human test to find wide application; it involved the application of a "standard patch" containing a test material to the skin of the arm or back. Iterations of the Schwartz-Peck Test included the Traub-Tusing-Spoor Method and the Brunner-Smiljanic Test and involved varying ranges of time for application of the patches[18]. In 1944 John Draize described methods to study skin irritation and local toxicity to skin and mucous membranes using several animal models and human testing. These tests involved exposing rabbits, dogs, white male guinea pigs, and humans to single and multiple dose applications of varying known sensitizing chemicals via rubberized cloth or surgical gauze that was bound to the area. Sensitization was evaluated by measuring epidermal abrasions using an arbitrary scoring system to indicate severity of erythema and edema. Draize recognized that while testing sensitizers in the skin of humans was ideal for accurate data, it posed risk of sensitizing

humans to compounds with which they could later come in contact. As such, Draize suggested it would be preferable to test compounds first on laboratory animals, and guinea pigs in particular[19]. The Draize test was recommended by the US. Food and Drug Administration (FDA) and was widely used for a period[20].

The first streamlined animal tests for screening skin sensitizers were developed in the 1960's. Buehler reiterated the limitations of using human subjects for testing skin sensitizers, specifically cost concerns, ambiguous results, chemical exposure, and the sample size necessary to ensure an adequate predictive value. The Buehler assay, or occluded (closed) patch test, was thus established in 1965 for testing on guinea pigs. Sensitization reactions were graded for intensity and duration of response as compared to the controls[21]. In 1966, Albert Kligman described the development of a streamline human assay for contact allergens to replace the preceding patch tests, called "the maximization test"[18,22-23]. Three years later, Kligman and Bertil Magnusson established the Guinea Pig Maximization Test (GPMT). The GPMT entailed initial intradermal injections with test agents and then topical application of the test agent to the same area one week later. Sensitization reactions were scored based on severity. The key factor in both the human maximization test and the GPMT was repeated occluded exposure, as the resulting reactions may distinguish between sensitizer (which will induce increasingly rapid or strong reactions due to T-cell memory) and irritant (which may not induce reaction, or reaction will fade rapidly). Kligman and Magnusson were able to compare the results from the GPMT to the human data from Kligman's human assay and determined the GPMT to be more accurate relative to human data than the Draize test[24]. The Buehler assay and the GPMT were the gold standards for twenty years.

The LLNA

In 1986 the preliminary stages of a new screening assay were set. Ears of BALB/c mice were "painted" with a test chemical or vehicle and the draining lymph nodes were excised and weighed. Lymph node cells were cultured and spontaneous proliferative capacity of these cells

was quantified by estimating [^{3}H] thymidine incorporation. The results were promising, and so it was determined that local lymph node responses could provide objective and quantitative data for screening skin sensitizers[25]. In 1989 the local lymph node assay (LLNA) was established, slightly modified from the original assay in 1986. In the established assay procedure, mice were exposed topically to chemicals on the dorsum of both ears and radioactive [^{3}H] thymidine was injected intravenously via the lateral tail vein. Lymphocyte proliferation was objectively quantified by thymidine incorporation in the draining lymph nodes[26]. The LLNA offered various advantages in comparison to the GPMT. The GPMT required subjective assessment of sensitization reactions and therefore the endpoint tended to be highly variable, while the LLNA provided quantitative data to describe sensitization reactions. Fold-difference in this data could also be used to estimate the sensitizer potency. Potency estimation was determined by the EC3, the effective chemical concentration required to induce a stimulation index (SI) of three relative to vehicle-treated controls. Higher EC3 values were associated with weaker sensitizers, while lower EC3 values correlated well with sensitizers known to be potent in humans. In particular, sensitizers with EC3 values $\leq$ 2% are categorized under GHS subcategory 1A, while > 2% are categorized as 1B. The LLNA also reduced the number of animals needed, decreased experimental time and costs, and improved animal welfare[27-28]. The LLNA was determined to be 86% accurate when compared to all guinea pig tests and was found to achieve similar accuracy (72% for both guinea pig data and LLNA data) to human data. The LLNA was subjected to extensive evaluation and was adopted as a screening test in combination with guinea pig tests by the Organization for Economic Cooperation and Development (OECD) in 1992[29]. The United States Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the European Center for the Validation of Alternative Methods (ECVAM) reviewed and validated the LLNA in 1999 and 2000 respectively, making it the preferred method for assessing skin sensitization potential[30-31]. In 2002 the LLNA was adopted as a stand-alone method, the new gold standard of skin sensitization testing, by the OECD[32].

1.3.2 ALTERNATIVE METHODS

Increasing ethical concerns regarding the use and quantity of animals for scientific research motivated focus on alternative methods. The European Center for the Validation of Alternative Methods (ECVAM) was established in 1991 in response to European Union (EU) Directive 86/609/EEC regarding protection of animals for scientific and other purposes. The goal of the ECVAM was to encourage research into the development and validation of alternative methods that could provide the same safety information gleaned from current animal tests using fewer animals, less painful procedures, or other means. The main goal of ECVAM was defined: "ECVAM will promote the scientific and regulatory acceptance of alternative methods which are of importance to the biosciences and which reduce, refine, or replace the use of laboratory animals"[33].

In the United States, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) was established as an ad hoc committee in 1994 under the NIH Revitalization Act of 1993 (Public Law 103-43, Sec. 205, pp. 25-27). Among their roles were to develop and validate assays and protocols, including alternative methods, that could reduce or eliminate the use of animals in safety testing, to establish criteria for the validation and regulatory acceptance of these alternative methods, and to recommend a process through which scientifically validated alternative methods could be accepted for regulatory use. In 2000 the ICCVAM was established as a permanent interagency committee of the National Institute of Environmental Health Sciences (Public Law 106-545, 42 USC 2851-5).

Improvements to the LLNA

To increase safety measures and reduce animal use, a number of modifications to the standard LLNA have been made and accepted. Two non-radioactive endpoints have been developed and approved to replace the use of [^{3}H] thymidine for lymphocyte quantification in order to eliminate occupational exposure to radioactivity and the issues related to radioactive waste. The first is the

LLNA: BrdU-ELISA (Enzyme-Linked Immunosorbent Assay). This assay uses non-radiolabelled 5-bromo-2-deoxyuridine (BrdU) in an ELISA-based test system to measure lymphocyte proliferation (OECD TG 442 B). The second assay is the LLNA: DA, which uses quantification of adenosine triphosphate (ATP), measured using bioluminescence, as an indicator of increased lymphocyte proliferation (OECD TG 442 A)[34]. However, there are limitations to these modifications. Although both of these assays provide quantitative data that may be used for dose-response assessment, the results may not be directly compared to the EC3 values obtained from the standard LLNA. The thresholds to determine sensitization, or SI, for these assays are lower than the SI for the standard LLNA. A limitation specifically of The LLNA: DA is the inability to test certain chemicals that may affect ATP levels and thus the results of the assay[27]. Furthermore, neither of these modified assays serve to reduce the number of animals used or introduce a less painful procedure.

The reduced LLNA (rLLNA) was introduced to reduce the number of animals used in the assessment of skin sensitization potential of chemicals. The rLLNA uses fewer animals than the standard LLNA to provide a simple yes/no result when there is a regulatory need to confirm a negative prediction of skin sensitization potential, but no need for dose-response assessment. This assay can reduce the number of animals needed for each test by 40% compared to the standard LLNA, and was approved by the OECD and recommended by ICCVAM and ECVAM when appropriate[35-37]. However, clear justification must be used must be provided before using this method, as it does not provide dose-response or potency data that could be used for chemical risk assessment.

Limitations of In Vivo (Animal) Testing

No predictive toxicology test will ever be absolutely accurate. Both the LLNA and guinea pig tests are found to be only 72% accurate to human data. Understanding the source of these limitations is critical. The first aspect to consider is the classification decision threshold used in

the guinea pig tests. According to the Buehler assay, if ≥15% of the guinea pigs tested test positive (by their subjective criteria), the chemical is classified as a potential sensitizer. If ≥30% of guinea pigs test positive in the GPMT (also by subjective criteria), the chemical is classified as a potential sensitizer. These thresholds seek to identify substances with "significant" skin sensitization potential, but inevitably leave room for misclassification of weaker sensitizers. The guinea pig tests are particularly susceptible to false positives and negatives that hover around those thresholds[38]. Induced false-positives were also found to result from hyperirritability due to the assay method, which involves "maximizing" exposure and enhancing allergenicity, and the initial lack of a proper control[39]. Inconsistencies between human maximization test data and the LLNA have been documented[40]. These inconsistencies may be attributed to interspecies differences. For example, rodents have been shown to demonstrate increased skin penetration of chemicals to humans[41]. This could lead to increased false positives. Additionally, lymphocyte proliferation has been shown to be influenced by vehicle selection. Several of the vehicles recommended by the OECD have been shown to augment the LLNA response to certain chemicals or cause contact allergy in mice. The LLNA also does not distinguish between specific types of hypersensitivity; specifically, it cannot distinguish between ACD, a Th1-type immune response, and sensitization of the respiratory tract, a Th2-type immune response[27]. These limitations among others and the related ethical concerns contributed to the ban of animal testing on cosmetic products and ingredients in the European Union and the marketing of such products effective as of 2013[42]. Tremendous efforts around the world have been made to introduce accurate alternative methods and eliminate or reduce animal testing of cosmetic products. Most alternative methods attempt to predict one or more of the key events of sensitization: (1) protein binding, (2) keratinocyte activation, (3) dendritic cell activation, and (4) T cell activation. These four events are known together as the adverse outcome pathway (AOP) of ACD.

In Silico Approaches

Landsteiner and Jacobs were the first to notice a correlation between chemical reactivity and skin sensitization[8]. The identification of electrophilic features in those chemicals was the basis of the studies investigating associations between structure and sensitization response. Dupuis and Benezra introduced the concept of structure activity relationships (SARs) in contact allergy and demonstrated that the structural requirements for haptens are highly specific[43]. These structural relationships were based off datasets from the LLNA and guinea pig tests. Efforts were also made to identify "structural alerts" and common reaction mechanisms. The original SARs gave only qualitative information, but it was useful to identify potential sensitizers as a first stage. A common approach to SARs is the Relative Alkylation Index (RAI), a mathematical model derived by Roberts and Williams. The RAI model analyzes electrophilicity and hydrophobicity parameters as well as chemical dose to establish SARs for skin sensitizers[44]. Deductive Estimation of Risk from Existing Knowledge (DEREK) for Windows is a knowledge-based expert system that uses prior knowledge of structure-toxicity relationships to create structural alerts for molecules with similar structures. In the case of skin sensitizers, each alert describes a structural feature that generally has the potential for electrophilic binding to skin proteins[45]. The DEREK system evolves as new knowledge is acquired. DEREK for Windows is now called Derek Nexus (v.3.0.1). This updated, rule-based SAR generates toxicity prediction and assigns confidence terms to the prediction (probable, plausible, improbable, equivocal, and nothing to report)[46]. Toxtree is another SAR software that groups chemicals according to their mechanistic category[47].

In 1989 it was recommended at an OECD workshop that an attempt be made to evaluate the predictive power of Quantitative Structure Activity Relationships (QSARs). The overarching goal was to apply statistical methods to sets of biological data used in SARs to predict a yes/no sensitizing outcome. The initial results reported for use of QSAR to predict skin sensitization were not particularly promising; there was high incidence of false negatives. However, the

concurrence of positive QSAR predictions with positive animal results offered the possibility to replace animal testing of chemicals predicted to be skin sensitizers with QSAR predictions, and so it was pursued further[48]. The OECD QSAR Toolbox v.3.4 is freely available online at qsartoolbox.org. It contains a database of existing skin sensitization data (mainly from the LLNA and GPMT) and predicts skin sensitization potential by read-across. The software includes different metabolic and transformation simulators, including a skin metabolism simulator[47,49]. Toxicity Prediction Komputer-Assisted Technology (TOPKAT) and Computer Automated Structure Evolution program (CASE) use a statistical/empirical approach to identify structural fragments associated with sensitization independently of action mechanism and develop a probability value of a compound to be a sensitizer or a non-sensitizer. VEGA is another QSAR software that uses the adaptive fuzzy partitioning (AFP) algorithm based on 8 descriptors in a multivariate statistical model. The output of the model is "active" or "inactive" for sensitizer prediction. TIssue METabolism Simulator – Skin Sensitization (TIMES-SS) is a hybrid expert system that encodes structure-toxicity and structure-skin metabolism relationships through a number of transformations. TIMES-SS offers an approach to predict skin sensitization potential of chemicals while taking chemical metabolism into account with a metabolism simulator. This simulator uses a hierarchal list of spontaneous, enzyme-mediated, and protein-binding reactions that the chemicals in question can undergo[47,49-50].

In Chemico Approaches (Protein Binding Assays)

In 2004 a new screening assay emerged that capitalized on the binding of sensitizing chemicals with skin proteins (step 1 in the AOP). Instead of computational methods to predict peptide reactivity, this approach (the Direct Peptide Reactivity Assay, or DPRA) uses binding of allergens to peptides commonly found in the skin to predict hapten-protein complexes in the skin that could induce sensitization. In particular, the initial assay used nucleophilic peptides lysine, cysteine, and histidine, and found cysteine to be the most reactive with the chemicals tested[51]. A

10% peptide depletion cut-off was used to categorize chemicals and sensitizers or non-sensitizers. Later iterations of the assay involved the inclusion of horseradish peroxidase and hydrogen peroxide for enzymatic activation of pro-haptens (the Peroxidase Peptide Reactivity Assay, or PPRA) and then the re-incorporation of lysine to the PPRA for detection of pre- and pro-haptens[52-53].

In Vitro Approaches

Several investigative groups have made efforts to target the second, third, and fourth steps of the AOP via *in vitro* approaches to sensitizer prediction. They use different cell types and measure a spectrum of biomarkers to screen skin sensitizers (**Table 1.1**).

A number of assays target the second stage in the AOP, keratinocyte activation. The metabolic competency of skin is well-documented; specifically, at least 36 xenobiotic-metabolizing enzyme proteins can be detected by proteomic profiling in the skin. The levels in the skin are 4-10 times lower than in the liver. Keratinocytes have been shown to express both phase I and phase II metabolizing enzymes. For a more comprehensive review see Dumont, et al[49]. Enzymes in the skin are able to metabolize both endogenous and exogenous compounds with the presumed goal of detoxifying potentially reactive chemicals; however, this can also lead to the activation of the compound (as in the case of pro-haptens). The KeratinoSens™ and LuSens assays both use antioxidant response element (ARE) [Luciferase] reporter elements in keratinocyte cell lines to screen skin sensitizers. Studies have shown that the Nrf2-Keap1 pathway plays an important role in skin sensitization. Under normal physiological conditions Nrf2 is complexed with Keap1. The Keap1 protein contains highly reactive cysteine residues. In response to covalent modification by stressors (i.e. allergens), Keap1 releases Nrf2. The free Nrf2 translocates to the nucleus, complexes with other molecules, and then binds to the ARE in the promoter region, thus initiating transcription of the downstream genes. As such, expression of luciferase downstream from ARE indicates binding of a skin sensitizer and can be used as a predictive metric of skin sensitization[54-

[57]. EVCAM issued their recommendations for KeratinoSens™ in 2013, describing the advantages and limitations of the assay and recommended its use in conjunction with other assays[58]. Another keratinocyte-based assay is the NCTC 2544 IL-18 assay, which uses secretion of IL-18 as a predictive metric of skin sensitization. Keratinocytes have been shown to express IL-18 mRNA and protein. IL-18 is a member of the IL-1 cytokine family and is a potent inducer of IFN-γ in activated T-cells. IL-18 has been shown to favor Th1 immune responses by promoting secretion of pro-inflammatory mediators like TNF-α, IL-8, and IFN-γ and to play a key role in the induction of ACD. It was therefore chosen as a metric to predict sensitization[59-61].

Dendritic cell (DC)-based assays are among the most promising approaches to distinguish sensitizers from non-sensitizers. Primary cells and a number of DC cell lines have been used to model sensitization outcomes. Monocyte-derived DCs (moDCs) from human peripheral blood have been studied as candidates for *in vitro* assay development. In particular, up-regulation of CD86 expression and CXCL8 (IL-8) secretion have been measured as predictive metrics of sensitization in moDCs[62-63]. moDCs have also been used to study T-cell mediated immune response in the skin. In particular, Saalbach et al demonstrated that dermal fibroblasts induced maturation of moDCs and may participate in the regulation of inflammation and immune responses in the skin[64]. Other groups have attempted differentiating monocytes to DCs expressing Langerin (moLCs) to be more similar to Langerhans cells. However, these assays are limited by availability of fresh blood and the complex, time-consuming protocols to establish primary DC cultures. Furthermore, donor variability makes it quite difficult to standardize moDC prediction models[65]. Using DC-like cell lines that are readily available and can be grown in a relatively short time under standardized conditions in place of primary cells addresses the issues of donor variation and availability. It is critical that these DC-like lines exhibit comparable phenotypic changes to blood-derived DCs upon exposure to sensitizers to be considered suitable candidates for *in vitro* screening assays. Among the most commonly utilized DC-like lines for *in vitro* screening assays are the human monocytic leukemia line THP-1, the human histiocytic

lymphoma line U-937, and the CD34+ human acute myeloid leukemia line MUTZ-3. The U-937 line was used in the development of the U-SENS™ assay, previously known as the Myeloid U937 Skin Sensitization Test (MUSST). U-SENS™ measures dendritic cell activation following exposure to sensitizers by expression of the co-stimulatory molecule CD86[66-68]. The THP-1 DC-like line has been used in a number of studies and in the development of established assays. Among the most prominent assays to use the THP-1 line is the human Cell Line Activation Test (h-CLAT). Similar to U-SENS™, h-CLAT measures up-regulation of the co-stimulatory markers CD86 and CD54 in response to skin sensitizers[69]. The OECD and EVCAM recommended h-CLAT as a validated assay for screening skin sensitizers together with complementary information in 2014 and 2015, respectively[70-71]. A stable THP1-derived IL-8 reporter line was established for the IL-8 Luc assay for sensitizer screening. IL-8 is reported to be important in the DC activation step in the AOP for skin sensitization[72-73]. The IL-8 Luc assay has been validated by the OECD as part of an integrated testing strategy for the predictive identification of skin sensitizers[74]. THP-1s have been used to study additional biomarkers of skin sensitization, including cell-surface thiols and reactive oxygen species (ROS)[75-76]. THP-1s have also been used in conjunction with other cells or agents to improve pre- and pro-hapten prediction, as THP-1s alone have been shown to express relatively low levels of metabolic enzymes compared to human skin. In particular, THP-1s have been co-incubated with rat liver microsomes (S9) for increased bioactivation of pro-haptens with moderate success, though S9 was found to slightly decrease surface marker up-regulation[77]. Furthermore, S9 is not entirely representative of the skin metabolic environment, as the levels of metabolic enzyme expression in the skin have been found to be 300-fold lower than those of the liver[49]. THP-1s were also used in a co-culture study with the HaCaT keratinocyte cell line. In that study, changes in CD86, CD40, and CD54 expression were used to predict sensitization. The addition of keratinocytes to the culture system and cross talk between the two cell types was found to improve pre- and pro-hapten detection, but this assay was not pursued further[78].

The MUTZ-3 line has been extensively studied and established as a suitable model of dendritic cells. When compared to other DC-like lines, MUTZ were found to be superior in activating resting T-cells. They were also found to have a similar expression profile to moDCs in both their immature and mature forms. MUTZ-3 cells can be activated with inflammatory cytokines to acquire a phenotype similar to Langerhans-like DCs[79]. MUTZ have been used in a number of sensitization studies with a variety of biomarkers, ranging from surface markers CD86 and CD54 to mRNA levels of inflammatory mediators to IL-8 secretion[80-82]. It was shown that analysis of IL-8 secretion proved a more successful method of distinguishing sensitizers from non-sensitizers than CD86 expression in MUTZ-derived Langerhans cells (MUTZ-LCs). It was further demonstrated that MUTZ-LCs migrate towards receptors involved in homing to the local lymph node (e.g. CXCL12) in response to skin sensitizers and migrate towards CLL5, a skin-homing receptor, in response to non-sensitizers[82]. The GARD® assay is a MUTZ-based assay currently undergoing validation studies by the OECD and ECVAM. GARD uses a 200-gene signature measured by genome expression array to distinguish between sensitizers and non-sensitizers. A support vector machine was used to validate the predictive power of the signature. Some of the pathways likely invoked by the molecules in the signature include Nrf-2 mediated oxidative response, xenobiotic metabolism signaling, protein ubiquination, LPS/IL-1 mediated inhibition of RXR function, aryl hydrocarbon receptor signaling, and protein kinase A signaling. These pathways are known to take part in xenobiotic-provoked reactions[83-84]. A summary of the *in vitro* assays recommended by the OECD or undergoing validation can be found in **Table 1.1**.

Assay/Group	Cell(s)	Cell Type	Metric	Description
GARD®[83]	MUTZ-3	Dendritic cell	200 gene signature	Analyzes differential gene expression in response to sensitizers and non-sensitizers
h-CLAT[69]	THP-1	Dendritic cell	CD86, CD54	Measures augmentation of CD86 and CD54 in response to sensitizers
IL-8 Luc[72]	THP-G8	Dendritic cell	IL-8 reporter	Measures IL-8 expression in response to sensitizers
KeratinoSens™[,54]	HaCaT KC	Keratinocyte	ARE/luciferase	Measures activation of Nrf2/Keap1 pathway
LuSens[57]	LuSens™	Keratinocyte	ARE/luciferase	Measures activation of Nrf2/Keap1 pathway
NCTC IL-18[59]	NCTC	Keratinocyte	IL-18	Measures IL-18 secretion in response to sensitizers
U-SENS™ (MUSST)[67]	U937	Dendritic cell	CD86	Measures induction of CD86 expression

Table 1.1 Summary of *in vitro* assays under review or recommended by the OECD.

3D Skin Models

3D skin models offer the most physiologically relevant non-animal method of screening skin sensitizers. They offer the benefit of testing whether a particular chemical can penetrate the stratum corneum barrier into the epidermis. Furthermore, 3D models can be used to test chemicals that have low aqueous solubility, a considerable limitation in 2D *in vitro* models, which require the use of cellular medium. The three main classes of 3D skin models include excised skin, engineered epidermal models, and full-thickness skin equivalents. Freshly excised skin presents a number of limitations, including donor variability and relatively short shelf life, and thus will not be discussed further in this review[85]. Engineered epidermal models (also referred to as reconstructed human epidermal (RhE) models) using human keratinocytes can be used to target step (2) in the AOP of skin sensitization. EpiDerm™ and Episkin™ are both RhE models that have been validated for skin irritation and corrosion[86]. However, the inclusion of fibroblasts in the skin model may be important, as fibroblasts have been shown to induce maturation of DCs in the skin and to produce inflammatory mediators[64]. Fibroblasts and keratinocytes have also demonstrated a synergistic relationship in their roles in the AOP of skin sensitization[85]. Full-thickness skin equivalents incorporate both keratinocytes and fibroblasts and

therefore reflect more physiologically relevant conditions. Full-thickness models include Phenion™, EpidermFT™, TestSkin, AST-200, RealSkin, and Apigraf[49]. These models are generally readily available with minimal batch-to-batch variability. However, they are quite expensive, and therefore do not yet present a cost-effective method to screen the numbers of newly registered chemicals anticipated to require hazard identification and classification.

1.4 LIMITATIONS AND CHALLENGES

In Silico & In Chemico Approaches

Although DEREK and TOPKAT have been validated according to the OECD principles for QSARs, further improvements are still necessary. While the RAI approach is valuable in the generation of QSARs, it also limits prediction to congeneric sets of chemicals. Furthermore, many of the QSAR models are based on data from animal tests, and as such are constrained by the limitations of those assays[44]. Toxtree and the OECD QSAR toolbox highlight structural alerts related to skin sensitizing properties, but are not meant as stand-alone predictive tools. They also do not predict non-sensitizers; they specifically only alert the user to potential sensitizers[87]. Derek is limited in its prediction of metals and phenols[88]. One of the critical limitations of SAR and QSAR approaches is prediction of chemicals that require chemical conversion or metabolism. This introduces false negatives for inactive chemicals that form sensitizing agents *in vivo* or in air, i.e. pre- and pro-haptens[89]. TIMES-SS offers the most by way of simulation of metabolism, but is limited only to organic substances and excludes structures with unknown features from prediction, and thus does not offer a suitable stand-alone screening assay[87].

While the DPRA shows much promise in the prediction of skin sensitizers, it is particularly limited in its prediction of pre- and pro-haptens. The lack of an incorporated metabolic system in the initial assay prevented accurate prediction of naturally inert chemicals that require conversion prior to acting as a sensitizing agent. Modifying the DPRA to the PPRA by adding a peroxidase enzyme helped improve prediction of pro-haptens, but this assay has not been further developed

or validated[47]. Furthermore, the DPRA is not applicable for the testing of metal compounds, as those chemicals are known to react with proteins by different mechanisms[90]. Though *in silico* and *in chemico* models offer insight into a critical step in the sensitization process, more physiologically relevant assays should be considered in tandem for sensitizer prediction. A summary of the *in silico* and *in chemico* models and their current sensitizer prediction accuracy scores can be found in **Table 1.2**. These accuracy scores demonstrate a need for improvement. Low accuracy in prediction of pre- and pro-hapten sensitizers contributes to the overall reduction of their scores.

In Silico			
	Method	**Output**	**Accuracy OR Sensitivity/Specificity**
DEREK for Windows	SAR	Confidence terms	76[87]-78%[46]
TOPKAT	QSAR	Probability assessment	60/48%[87]-67/56%[47]
Vega VegaNic	QSAR	Active/Inactive	48%[46]
Toxtree	SAR	Qualitative	66%[46]
OECD QSAR Toolbox	QSAR	Structural alerts	60%[46]
TIMES-SS	QSAR	Significant, weak, NS	34-87%[50]
CASE	QSAR	Probability assessment	88/11%[47]
In Chemico			
DPRA	-	% peptide depletion	73[90]-89%[53]

Table 1.2 Partial list and summary of current *in silico* and *in chemico* methods for sensitizer prediction.

In Vitro Approaches

Though a number of *in vitro* methods have been validated or are undergoing validation studies for use as screening assays as part of an integrated testing strategy, these assays still present a number of limitations (summarized in **Table 1.3**). The KeratinoSens™ and LuSens assays are both limited to prediction of sensitizers that act through the Nrf2-Keap1 pathway. Sensitizers that do not react with cysteine peptides in the skin, such as phthalic anhydride, will not act through the Nrf2-Keap1 pathway, and therefore will not activate ARE-dependent gene activity and be detected as a false negative[91]. This limits the applicability of the keratinocyte ARE-luciferase-based assays. The GARD® assay, h-CLAT, IL-8 Luc, and U-SENS™ all use DC-like lines, addressing the third step in the AOP in skin sensitization. However, these single-cell assays lack the metabolic components present in human skin. Human skin functions as a protective barrier

against chemical exposure. Chemicals that pass that barrier are metabolized in the skin, which can lead to the activation of certain compounds (i.e. pro-haptens). Keratinocytes and dermal fibroblasts naturally express various CYP enzymes, including CYP1A1, CYP1B1, CYP2B6, CYP2E1, and CYP3A[12]. It has been further shown that enzyme activity can be induced in human skin[49]. Representing that metabolic activity in the skin sensitization process is critical to a more accurate prediction of skin sensitizers, and thus the lack of added metabolic components in the DC-based assays likely accounts for their lower pre- and pro-hapten prediction accuracy (**Table 1.3**). Another challenge facing the U-SENS™ and h-CLAT assays in particular is the use of CD86 as the metric for prediction of sensitization, as it has been shown that CD86 is not a reliable biomarker for detecting sensitizers[82].

With the exception of the GARD assay, all the *in vitro* assays validated or undergoing validation only investigate one facet or biomarker of the multi-step sensitization process. Indeed, this may account for the higher overall accuracy scores of the GARD assay; however, the lack of a metabolic component limits its prediction of pre- and pro-hapten sensitizers. It stands to reason that single-cell or single-metric assays are insufficient to screen all skin sensitizers. Though the field of alternative screening methods has expanded significantly over the last 20 years, there is room for improvement. In particular, a more physiologically relevant *in vitro* system combined with the use of a multi-metric biomarker signature could be more sensitive than a single cell assay or a single biomarker.

Assay	Cell	Cell Type	Metric	Accuracy	Pre-/Pro-Hapten Accuracy
GARD	MUTZ-3	Dendritic cell	200 gene signature	83[83]-98%[82]	75%[82]
h-CLAT	THP-1	Dendritic cell	CD86, CD54	85[69]-93%[92]	63%[92]
IL-8 Luc	THP-G8	Dendritic cell	IL-8 reporter	73[74]-88%[72]	75%[72]
KeratinoSens™	HaCaT	Keratinocyte	ARE	77-96%[55,58]	77%[91]
LuSens	LuSens™	Keratinocyte	ARE	83[56]-85%[57]	44%[56]
NCTC IL-18	NCTC 2544	Keratinocyte	IL-18	70[93]-97%[60]	Not enough data
U-SENS™	U937	Dendritic cell	CD86	77[68]-88%[67]	66%[68]

Table 1.3 List of current prominent *in vitro* screening methods and accuracy scores.

1.5 PREVIOUS WORK

One of the major challenges of many current screening assays is the prediction of pre- and pro-hapten sensitizers, as they do not incorporate a metabolic component to mimic the metabolic environment of human skin. Realskin, a reconstructed full-thickness skin model from EpiSkin™, was used in co-culture with MUTZ-3 derived Langerhans cells (MUTZ-LCs) to develop a culture platform that would be physiologically relevant and capable of metabolizing pre- and pro-haptens. This culture method was used to screen skin sensitizers, and the secretome from post-sensitized cultures was evaluated for 27 inflammatory cytokines and chemokines. Support vector machine (SVM) classification tools (a type of machine learning) were employed to determine the predictive capabilities of this culture system[94].

Machine learning involves the construction of algorithms that can learn from and make predictions on data using a "training" dataset and a "testing" set. In supervised machine learning, the algorithm is "trained" with input data assigned to a class or specific output and the goal is to determine the ability of the "test" data to adhere to the rule defined by the training data. A support vector machine is a supervised machine learning method used for classification and regression. Given data points assigned to one of two classes (e.g. sensitizer and non-sensitizer) to train the system, the goal of the algorithm is to determine to which class a "test" data point will belong. Each data point will be a p-dimensional vector, and the algorithm attempts to separate these vectors with some multi-dimensional hyperplane. Multiple hyperplanes could theoretically classify the data; the SVM attempts to choose a hyperplane that maximizes the separation, or margin, between the two classes.

The SVM was used to calculate the maximum margin distances for all 27 cytokines. These margin distances were ranked to identify the metrics that had the greatest distance of separation between non-sensitizers and sensitizers, where the top-ranking biomarkers show the most promise as metrics of prediction of sensitization. The SVM was also used as a classification model to determine accuracy of prediction. Model performance was assessed using k-fold cross validation

for each cytokine metric and for combined molecular signatures. SVM analysis with the Realskin and MUTZ-LC culture platform demonstrated 92% accuracy in prediction of sensitizers using the top 4 ranking biomarkers, IL-12, IL-9, VEGF, and IFN-γ. These studies demonstrated that a multi-metric approach with a physiologically relevant culture system has the potential to be more predictive of skin sensitizers[94]. However, this assay was found to be extremely costly and resource intensive. Therefore, a new goal was set to develop a 2D culture model that could mimic the environment of sensitization achieved with the Realskin/MUTZ model while maintaining comparable accuracy. To achieve this, a co-culture system of MUTZ-LCs, HaCaT keratinocytes (KCs), and primary dermal fibroblasts (FBs) was established. IL-8 was chosen as an initial test marker of sensitization based on precedence as a predictive biomarker of sensitization, and with MUTZ cells in particular, in the literature (72-73, 81-82). The initial results showed promise, and so this method was pursued (Serom Lee, unpublished work).

1.6 SUMMARY OF THESIS

The primary goal of this thesis work is to expand on a previously developed *in vitro* culture system and identify multiple screening metrics and methods using one high-throughput, cost-effective assay while establishing high accuracy and improved prediction of pre- and pro-haptens and sensitizer potency. The prevalence of allergic contact dermatitis and widespread use of potentially sensitizing agents in everyday cosmetic products necessitates accurate screening assays. Though animal testing is the current gold standard, limited accuracy, cost, rising ethical concerns with the use of animals in scientific research, and the recent EU ban of animal testing on cosmetic products and ingredients have motivated research of alternative screening methods. A number of *in silico*, *in chemico*, and *in vitro* alternatives have been developed, and some have been validated by the OECD, but these assays have particularly limited prediction of pre- and pro-hapten sensitizers, which require abiotic or metabolic activation respectively prior to inducing sensitization. Previous work using full-thickness skin equivalents and multi-metric prediction

using a support vector machine managed to overcome these limitations and offer high prediction accuracy, but engineered skin models are costly and low-throughput. As such, an *in vitro* co-culture system using MUTZ-3-derived Langerhans cells, HaCaT keratinocytes, and primary dermal fibroblasts was developed to provide a physiologically relevant sensitizer-screening assay that mimics the *in vivo* environment of sensitization. Initial studies measuring IL-8 secretion in this co-culture system showed promise, and so further development of this assay was pursued.

In chapter 2 we expand our studies on our developed *in vitro* co-culture system with a larger chemical panel and use computational tools to assess the statistical accuracy of prediction. We compared IL-8 secretion in the co-culture system to MUTZ-LCs alone and determined that up-regulation of IL-8 secretion could be used as a predictive metric of sensitization. To improve this prediction we utilized a 27-cytokine multi-plex assay and used our previously developed support vector machine to identify and rank the most predictive metrics in combination. We further employed classification trees to determine a statistical threshold to distinguish between sensitizer and non-sensitizer and sensitizer potencies. These methods identified the co-culture system to be superior to a single-cell assay in prediction of all sensitizer classes and potency. It was also found that a combination of predictive metrics achieves higher accuracy than one alone. To identify more predictive metrics, we deemed it necessary to break down the co-culture system and analyze how the different cell types contributed to the sensitization process.

In chapter 3 we study IL-8 secretion in response to skin sensitizers and irritants in the varying cell types used in our co-culture model in every mono-culture and co-culture combination possible. The motivation behind these studies was to identify the cells responsible or necessary for up-regulation of inflammatory cytokines. We found that keratinocytes and fibroblasts secrete little IL-8 and that IL-8 is only up-regulated in the presence of MUTZ-LCs. These findings will be used in future studies.

Chapter 4 includes a summary of the key findings of this thesis and their implications and the future directions of this work.

1.7 REFERENCES

1. Cashman, M. W., et al. (2012). "Contact dermatitis in the United States: epidemiology, economic impact, and workplace prevention." Dermatol Clin **30**(1): 87-98, viii.
2. Thyssen, J. P., et al. (2007). "The epidemiology of contact allergy in the general population--prevalence and main findings." Contact Dermatitis **57**(5): 287-299.
3. Peiser, M., et al. (2012). "Allergic contact dermatitis: epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. Current knowledge assembled at an international workshop at BfR, Germany." Cell Mol Life Sci **69**(5): 763-781.
4. Vocanson, M., et al. (2009). "Effector and regulatory mechanisms in allergic contact dermatitis." Allergy **64**(12): 1699-1714.
5. Enk, A. H. (1997). "Allergic contact dermatitis: understanding the immune response and potential for targeted therapy using cytokines." Mol Med Today **3**(10): 423-428.
6. Spiekstra, S. W., et al. (2005). "Induction of cytokine (interleukin-1alpha and tumor necrosis factor-alpha) and chemokine (CCL20, CCL27, and CXCL8) alarm signals after allergen and irritant exposure." Exp Dermatol **14**(2): 109-116.
7. Karlberg, A. T., et al. (2008). "Allergic contact dermatitis--formation, structural requirements, and reactivity of skin sensitizers." Chem Res Toxicol **21**(1): 53-69.
8. Landsteiner, K. and J. Jacobs (1935). "Studies on the Sensitization of Animals with Simple Chemical Compounds." J Exp Med **61**(5): 643-656.
9. Aleksic, M., et al. (2007). "Investigating protein haptenation mechanisms of skin sensitisers using human serum albumin as a model protein." Toxicol In Vitro **21**(4): 723-733.
10. Chipinda, I., et al. (2011). "Haptenation: chemical reactivity and protein binding." J Allergy (Cairo) **2011**: 839682.
11. Karlberg, A. T., et al. (2013). "Activation of non-sensitizing or low-sensitizing fragrance substances into potent sensitizers - prehaptens and prohaptens." Contact Dermatitis **69**(6): 323-334.
12. Bergstrom, M. A., et al. (2007). "A skin-like cytochrome P450 cocktail activates prohaptens to contact allergenic metabolites." J Invest Dermatol **127**(5): 1145-1153.
13. Kaplan, D. H., et al. (2012). "Early immune events in the induction of allergic contact dermatitis." Nat Rev Immunol **12**(2): 114-124.
14. Patlewicz, G., et al. (2016). "Can currently available non-animal methods detect pre- and pro-haptens relevant for skin sensitization?" Regul Toxicol Pharmacol **82**: 147-155.
15. Committee on the Design and Evaluation of Safer Chemical Substitutions: A Framework 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 Oct 29. Appendix D, Overview of the GHS Classification Scheme in Hazard Classification. Available from: https://www.ncbi.nlm.nih.gov/books/NBK253967/
16. Safford, R. J., et al. (2015). "Extension of the Dermal Sensitisation Threshold (DST) approach to incorporate chemicals classified as reactive." Regul Toxicol Pharmacol **72**(3): 694-701.
17. Geier, J., et al. (2015). "Fragrance mix I and II: results of breakdown tests." Flavour Fragr J **30**: 264-274.
18. Kligman, A. M. (1966). "The identification of contact allergens by human assay. I. A critique of standard methods." J Invest Dermatol **47**(5): 369-374.
19. Draize, J. H., et al. (1944). "Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes." J Pharm Exp Ther **82**: 377-390.
20. Draize, J. H. Dermal toxicity, p 46, "Appraisal of the Safety of Chemicals in Foods, Drugs, and Cosmetics." The Association of Food & Drug Officials of the United States, Texas State Dept. of Health, Austin, Texas. 1959.
21. Buehler, E. V. (1965). "Delayed Contact Hypersensitivity in the Guinea Pig." Arch Dermatol **91**: 171-177.

22. Kligman, A. M. (1966). "The identification of contact allergens by human assay. II. Factors influencing the induction and measurement of allergic contact dermatitis." J Invest Dermatol **47**(5): 375-392.
23. Kligman, A. M. (1966). "The identification of contact allergens by human assay. 3. The maximization test: a procedure for screening and rating contact sensitizers." J Invest Dermatol **47**(5): 393-409.
24. Magnusson, B. and A. M. Kligman (1969). "The identification of contact allergens by animal assay. The guinea pig maximization test." J Invest Dermatol **52**(3): 268-276.
25. Kimber, I., et al. (1986). "Development of a murine local lymph node assay for the determination of sensitization potential." Fd Chem Toxic **24**(6/7): 585-586.
26. Kimber, I., et al. (1989). "The murine local lymph node assay for identification of contact allergens: a preliminary evaluation of in situ measurement of lymphocyte proliferation." Contact Dermatitis **21**(4): 215-220.
27. Anderson, S. E., et al. (2011). "The LLNA: A Brief Review of Recent Advances and Limitations." J Allergy (Cairo) **2011**: 424203.
28. Corsini, E., et al. (2016). "Alternative Approach for Potency Assessment: In Vitro Methods." Cosmetics **3**(1): 7.
29. OECD, 1992. "OECD Guideline for the Testing of Chemicals No. 402: Acute Dermal Toxicity." Organization for Economic Cooperation and Development, Paris, France.
30. ICCVAM. "Interagency Coordinating Committee on the Validation of Alternative Methods. The Murine Local Lymph Node Assay: a test method for assessing the allergic contact dermatitis potential of chemicals/compounds. The results of an independent peer review evaluation coordinated by the ICCVAM and the NICEATM." National Institute of Environmental Health Sciences. NIH Publication no. 99-4494, https://ntp.niehs.nih.gov/
31. Balls, M., Hellsten, E., 2000. "Statement on the validity of the local lymph node assay for skin sensitization testing." ECVAM Joint Research Centre. European Commission, Ispra. ATLA 28, 366-367.
32. OECD, 2002. "Guideline for the Testing of Chemicals No. 429: Skin Sensitization: Local Lymph Node Assay." Organization for Economic Cooperation and Development, Paris, France.
33. Marafante, E., et al. (1994). "ECVAM: the European Centre for the Validation of Alternative Methods." Toxicol In Vitro **8**(4): 803-805.
34. OECD, 2010. "OECD Guideline for the Testing of Chemicals No. 442B: Skin Sensitization: Local Lymph Node Assay: BrdU-ELISA." Organization for Economic Cooperation and Development, Paris, France.
35. OECD, 2010. "OECD Testing Guideline 429, 429A, 429B. Skin Sensitization: Local Lymph Node Assay. Updated Guideline, Adopted 22th July 2010." Organization for Economic Cooperation and Development, Paris, France.
36. ICCVAM, "Interagency Coordinating Committee on the Validation of Alternative Methods. ICCVAM Test Method Evaluation Report. The Reduced Murine Local Lymph Node Assay: An Alternative Test Method. Using Fewer Animals to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products." National Institute of Environmental Health Sciences, NIH Publication No. 09-6439, https://ntp.niehs.nih.gov/
37. ESAC, 2007. "Statement on the Reduced Local Lymph Node Assay (rLLNA)." European Commission Directorate General, Joint Research Centre, Institute for Health and Consumer Protection, European Centre for the Validation of Alternative Methods, April 2007.
38. Basketter, D. A. and I. Kimber (2010). "Skin sensitization, false positives and false negatives: experience with guinea pig assays." J Appl Toxicol **30**(5): 381-386.
39. Kligman, A. M. and D. A. Basketter (1995). "A critical commentary and updating of the guinea pig maximization test." Contact Dermatitis **32**(3): 129-134.
40. Zaghi, D. and H. I. Maibach (2009). "The local lymph node assay compared with the human maximization test as an indicator of allergic potency in humans using patch test clinic populations." Cutan Ocul Toxicol **28**(2): 61-64.
41. Garnett, A., et al. (1994). "Percutaneous absorption of benzyl acetate through rat skin in vitro. 3. A comparison with human skin." Food Chem Toxicol **32**(11): 1061-1065.

42. EU, Directive 2003/15/EC of the European Parliament and of the Council of 27 February 2003 amending Council directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products, in Off. J. Eur. Union 2003. P. 26-35.
43. Dupuis, G., Benezra, C. Allergic Contact Dermatitis to Simple Chemicals. A Molecular Approach. Marcel Dekker, New York, 1982.
44. Smith Pease, C. K., et al. (2003). "Contact allergy: the role of skin chemistry and metabolism." Clin Exp Dermatol **28**(2): 177-183.
45. Thyssen, J. P., et al. (2012). "The critical review of methodologies and approaches to assess the inherent skin sensitization potential (skin allergies) of chemicals. Part I." Contact Dermatitis **66 Suppl 1**: 11-24.
46. Verheyen, G. R., et al. (2017). "Evaluation of in silico tools to predict the skin sensitization potential of chemicals." SAR QSAR Environ Res **28**(1): 59-73.
47. Ezendam, J., et al. (2016). "State of the art in non-animal approaches for skin sensitization testing: from individual test methods towards testing strategies." Arch Toxicol **90**(12): 2861-2883.
48. EPA and OECD, 1994. "US EPA/EC Joint Project on the Evaluation of (Quantitative) Structure Activity Relationships," Environment Monograph No. 88, Paris, France.
49. Dumont, C., et al. 2015. "Review of the availability of in vitro and in silico methods for assessing dermal bioavailability." Appl In Vitro Toxicol. **1**(2): 147-164.
50. Dimitrov, S. D., et al. (2005). "Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates." Int J Toxicol **24**(4): 189-204.
51. Gerberick, G. F., et al. (2004). "Development of a peptide reactivity assay for screening contact allergens." Toxicol Sci **81**(2): 332-343.
52. Gerberick, G. F., et al. (2009). "Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system." Toxicol Sci **112**(1): 164-174.
53. Troutman, J. A., et al. (2011). "The incorporation of lysine into the peroxidase peptide reactivity assay for skin sensitization assessments." Toxicol Sci **122**(2): 422-436.
54. Emter, R., et al. (2010). "Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro." Toxicol Appl Pharmacol **245**(3): 281-290.
55. Natsch, A., et al. (2011). "The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict sensitizers in vitro: Results of a ring-study in five laboratories." Toxicol In Vitro. **25**(3): 733-744.
56. Ramirez, T., et al. (2014). "LuSens: a keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification." Toxicol In Vitro **28**(8): 1482-1497.
57. Ramirez, T., et al. (2016). "Intra- and inter-laboratory reproducibility and accuracy of the LuSens assay: A reporter gene-cell line to detect keratinocyte activation by skin sensitizers." Toxicol In Vitro **32**: 278-286.
58. ECVAM, 2013. "EUR 26383 – draft EURL ECVAM recommendation on the KeratinoSens™ assay for skin sensitization testing." JRC Scientific and Policy Reports, pp. 35.
59. Corsini, E., et al. (2009). "Use of IL-18 production in a human keratinocyte cell line to discriminate contact sensitizers from irritants and low molecular weight respiratory allergens." Toxicol In Vitro **23**(5): 789-796.
60. Corsini, E., et al. (2013). "NCTC 2544 and IL-18 production: a tool for the identification of contact allergens." Toxicol In Vitro **27**(3): 1127-1134.
61. Galbiati, V., et al. (2011). "Further development of the NCTC 2544 IL-18 assay to identify in vitro contact allergens." Toxicol In Vitro **25**(3): 724-732.
62. Toebak, M. J., et al. (2006). "CXCL8 secretion by dendritic cells predicts contact allergens from irritants." Toxicol In Vitro **20**(1): 117-124.
63. Reuter, H., et al. (2011). "In vitro detection of contact allergens: development of an optimized protocol using human peripheral blood monocyte-derived dendritic cells." Toxicol In Vitro **25**(1): 315-323.
64. Saalbach, A., et al. (2007). "Dermal fibroblasts induce maturation of dendritic cells." J Immunol **178**(8): 4966-4974.

65. dos Santos, G. G., et al. (2009). "Progress on the development of human in vitro dendritic cell based assays for assessment of the sensitizing potential of a compound." Toxicol Appl Pharmacol **236**(3): 372-382.
66. Python, F., et al. (2007). "Assessment of the U937 cell line for the detection of contact allergens." Toxicol Appl Pharmacol **220**(2): 113-124.
67. Piroird, C., et al. (2015). "The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization." Toxicol In Vitro **29**(5): 901-916.
68. Ovigne, J.M., et al. "The Myeloid U937 Skin Sensitization Test (MUSST) for the prediction of skin sensitization potential." [Poster] 2011. Available from http://wc8.loreal.com/medias/4-themeI/THE%20MYELOID%20U937%20SKIN%20SENSITIZATION.pdf.
69. Nukada, Y., et al. (2012). "Prediction of skin sensitization potency of chemicals by human Cell Line Activation Test (h-CLAT) and an attempt at classifying skin sensitization potency." Toxicol In Vitro **26**(7): 1150-1160.
70. OECD, 2014. "OECD Guideline for the Testing of Chemicals No. 442E: In Vitro Skin Sensitization: human Cell Line Activation Test (h-CLAT)." Organization for Economic Cooperation and Development, Paris, France.
71. ECVAM, 2015. "EURL ECVAM recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitization testing." JRC Scientific and Policy Reports. European Commission Directorate General, Joint Research Centre, Institute for Health and Consumer Protection, European Centre for the Validation of Alternative Methods, February 2015.
72. Takahashi, T., et al. (2011). "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.
73. Kimura, Y., et al. (2015). "Optimization of the IL-8 Luc assay as an in vitro test for skin sensitization." Toxicol In Vitro **29**(7): 1816-1830.
74. OECD, 2016. "Validation report for the international validation study on the IL-8 Luc assay as a test evaluating the skin sensitizing potential of chemicals conducted by the IL-8 Luc Assay." Validation Management Team, December 2016. Available from http://www.oecd.org/env/ehs/testing/160405-validation-report_REV.pdf.
75. Suzuki, M., et al. (2009). "Evaluation of changes of cell-surface thiols as a new biomarker for in vitro sensitization test." Toxicol In Vitro **23**(4): 687-696.
76. Saito, K., et al. (2013). "Development of an in vitro skin sensitization test based on ROS production in THP-1 cells." Toxicol In Vitro **27**(2): 857-863.
77. Chipinda, I., et al. (2011). "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.
78. Hennen, J., et al. (2011). "Cross talk between keratinocytes and dendritic cells: impact on the prediction of sensitization." Toxicol Sci **123**(2): 501-510.
79. Larsson, K., et al. (2006). "Functional and transcriptional profiling of MUTZ-3, a myeloid cell line acting as a model for dendritic cells." Immunology **117**(2): 156-166.
80. Azam, P., et al. (2006). "The cytokine-dependent MUTZ-3 cell line as an in vitro model for the screening of contact sensitizers." Toxicol Appl Pharmacol **212**(1): 14-23.
81. Nelissen, I., et al. (2009). "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.
82. Ouwehand, K., et al. (2010). "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.
83. 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.
84. Johansson, H., et al. (2017). "Evaluation of the GARD assay in a blind Cosmetics Europe study." ALTEX.
85. Sharma, N. S., et al. (2012). "Perspectives on Non-Animal Alternatives for Assessing Sensitization Potential in Allergic Contact Dermatitis." Cell Mol Bioeng **5**(1): 52-72.

86. Speilmann, H. (2007). “The ECVAM international validation study on in vitro tests for acute skin irritation: report on the validity of the EPISKIN and EpiDerm assays and on the Skin Integrity Function Test.” Altern Lab Anim **35**: 559-601.
87. Teubner, W., et al. (2013). "Computer models versus reality: how well do in silico models currently predict the sensitization potential of a substance." Regul Toxicol Pharmacol **67**(3): 468-485.
88. Canipa, S. J., et al. (2017). "A quantitative in silico model for predicting skin sensitization using a nearest neighbours approach within expert-derived structure-activity alert spaces." J Appl Toxicol.
89. Ashby, J., et al. (1995). "Structure activity relationships in skin sensitization using the murine local lymph node assay." Toxicology **103**(3): 177-194.
90. Takenouchi, O., et al. (2015). "Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals." J Appl Toxicol **35**(11): 1318-1332.
91. Natsch, A. and R. Emter (2008). "Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitization potential of chemicals." Toxicol Sci **102**(1): 110-119.
92. Sakaguchi, H., et al. (2009). “The relationship between CD86/CD54 expression and THP-1 cell viability in an in vitro skin sensitization test – human cell line activation test (h-CLAT).” Cell Biol Toxicol. **25**(2): p. 109-26.
93. Guyard-Nicodeme, M., et al. (2015). "Development of a multiparametric in vitro model of skin sensitization." J Appl Toxicol **35**(1): 48-58.
94. Lee, S., et al. (2014). "Predicting full thickness skin sensitization using a support vector machine." Toxicol In Vitro **28**(8): 1413-1423.

CHAPTER 2: CXCL8 PRODUCTION IN CO-CULTURE OF KERATINOCYTES, DERMAL FIBROBLASTS, AND LANGERHANS CELLS TO DISTINGUISH CONTACT SENSITIZER AND CLASSIFY POTENCY

Note: Sections 2.2 and 2.3 are adapted from the following manuscript to be submitted for publication to *Toxicology and Applied Pharmacology*:

Serom Lee, **Talia Greenstein**, Lingting Shi, Tim Maguire, Rene Schloss, Martin Yarmush

CXCL8 Production in Co-culture of Keratinocytes, Dermal Fibroblasts, and Langerhans Cells to Distinguish Contact Sensitizer and Classify Potency

2.1 INTRODUCTION

Allergic Contact Dermatitis (ACD) is one of the most prevalent dermatoses, impacting approximately 15-20% of the general population[1]. ACD is a delayed hypersensitivity reaction that prompts an adaptive immune response mediated by antigen-specific T-cells, and is 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 may 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]. While a number of alternative methods have been established and validated for use as part of an integrated testing strategy, these assays have particularly limited accuracy in prediction of pre- and pro-hapten sensitizers. Furthermore, the majority of these assays use a

single metric to predict sensitization, and few make attempts to predict sensitizer potency (see chapter 1 for details). To address these limitations, we previously developed a co-culture system of MUTZ-3-derived Langerhans cells (MUTZ-LCs), HaCaT keratinocytes (KCs), and primary dermal fibroblasts (FBs) to screen skin sensitizers. FBs and KCs provide the metabolic components necessary to activate pre- and pro-haptens and work synergistically to promote the immune response[7]. IL-8 was used as a starting biomarker of sensitization, with the intention of expanding to multi-metric analysis.

Interleukin-8, also called IL-8 or CXCL8, is a chemotactic factor produced by a wide variety of cells, including stimulated monocytes, macrophages, and non-leukocytic cells including epithelial and endothelial cells, lymphocytes, fibroblasts, and keratinocytes[8]. It has been shown to activate and recruit neutrophils to inflammatory sites as well as T lymphocytes[9]. Furthermore, delayed-type hypersensitivity has been shown to depend on IL-8; IL-8 acts as a mediator of leukocyte chemotaxis and activation during T-cell mediated immune responses[10]. As such, it has been extensively studied as a biomarker of skin sensitization. Skin sensitizers enhanced IL-8 mRNA expression in moDCs and MUTZ-LCs[11-12]. The IL-8 Luc assay also quantifies IL-8 expression by a luciferase IL-8 reporter in THP-1 cells with success[13]. IL-8 secretion was found to uniformly increase after exposure to sensitizer but not non-sensitizers in MUTZ, as well[14]. The precedence and success of using IL-8 as a metric of predicting sensitization established its promise as an initial biomarker for our co-culture system. Previous work demonstrating the improved prediction accuracy of a panel of biomarkers in a full-thickness skin model using a support vector machine (SVM) and feature selection motivated similar such studies with the co-culture system[15]. In addition, classification trees have been used to improve skin sensitization hazard prediction with combined data from multiple screening methods[16]. We applied classification tree learning to our data to identify statistical thresholds to distinguish sensitizer and non-sensitizer and classify potency.

2.2 MATERIALS AND METHODS

Cell Culture

The HaCaT keratinocyte (KC) cell line was a donation from Dr. Bozena Michniak-Kohn. Human primary dermal fibroblasts (FB) were a donation from Dr. Francois Berthiaume. Both HaCaT KCs and FBs 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.

The 5637 human bladder carcinoma line was purchased from ATCC (Manassas, VA). The 5637 cell line was maintained in RPMI medium supplemented with 10% FBS, 2% L-glutamine, and 1% penicillin-streptomycin at 37°C and 5% CO_2. To collect conditioned media, cells were seeded at $19x10^6$ in 38 mL media in a T225 tissue culture flask and allowed to grow 48 hours to confluence. Media was changed and collected 42 hours later and stored at -80°C. This conditioned medium 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). The MUTZ-3 cell line was maintained in alpha-MEM medium with Glutamax, ribonucleosides, and deoxyribonucleosides (Invitrogen) supplemented with 20% heat-inactivated FBS, 10% 5637 conditioned medium, 1% penicillin-streptomycin, and 50 µM 2-mercaptoethanol (complete media) at 37°C and 5% CO_2. Media was changed every other day and the cells were split on day five of culture. To differentiate the MUTZ-3s to Langerhan-like cells (MUTZ-LCs), cells were seeded at 100,000 cells/mL and cultured for 7 days in complete media with 2.5 ng/mL TNF-α, 10 ng/mL TGF-β1, and 100 ng/mL GM-CSF. On days 2 and 5 of culture, media was changed and fresh cytokines were added at half those concentrations. Differentiated MUTZ-LCs utilized in experiments were maintained in MUTZ-3 media described above without the 5637 conditioned medium. This medium will be described as the Maturation Medium in future sections.

Chemicals and Reagents

Test chemicals included both non-sensitizers and known skin sensitizers of every class and potency (**Table 2.1**). Non-sensitizing chemicals (as classified by the LLNA) included the vehicle (0.1% dimethylsulfoxide), isopropanol (2PR), xylene (XYL), and four skin irritants: lactic acid (LA), salicylic acid (SA), sodium dodecyl sulfate (SDS), and vanillin (VL). A panel of 12 sensitizers of varying classes and potencies outlined in Table 2.1 were evaluated. This panel included cinnamic alcohol (CA), eugenol (EU), geraniol (GER), cinnamaldehyde (CLD), isoeugenol (IE), 2-methoxy-4-methylphenol (MMP), resorcinol (RC), 2-aminophenol (2AP), 2-4-dinitrochlorobenzene (DNCB), hydroquinone (HQ), p-benzoquinone (pBQ), and p-phenylenediamine (PPD). All chemicals were purchased through Sigma-Aldrich and prepared in dimethylsulfoxide (DMSO) and serially diluted in maturation media where the final concentration of DMSO in cell culture during treatment was ~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).

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

Table 2.1 Panel of chemicals evaluated.

Co-culture of HaCaT Keratinocytes, Dermal Fibroblasts, and MUTZ-3 Langerhans Cells

HaCaT KCs and human dermal FBs were plated at $1.25x10^4$ 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 $2.5x10^4$ 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. Duplicate plates of both the co-culture and mono-culture conditions were prepared where one plate was utilized to assess viability and the other was used to collect supernatant. 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 (**Figure 2.1**).

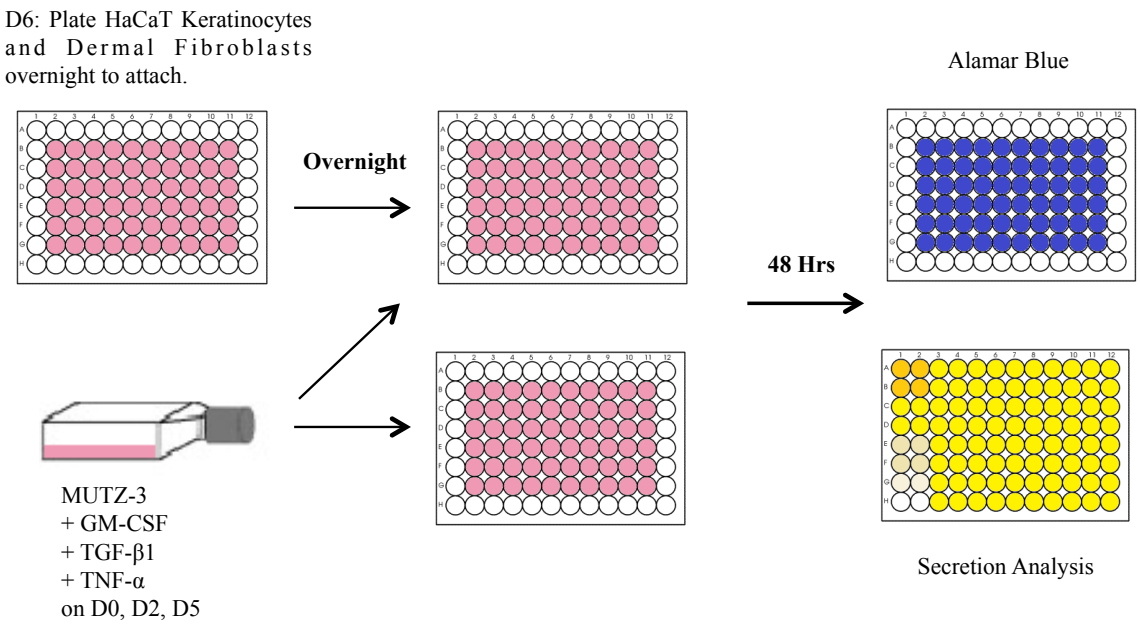

Figure 2.1 Experimental schematic for co-culture of MUTZ-LCs, HaCaT KCs, and dermal FBs.

Viability

Viability of both culture platforms was analyzed using the Alamar Blue™ assay, which measures reduction of resazurin to resorufin and thus is an indicator of the reducing environment of the living cell. The Alamar Blue assay was performed according to the manufacturer's protocol and reduction was 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:

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

Cytokine Secretion

Supernatant collected from both co-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-1β, 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. Outliers attributed to human or mechanical error (i.e. faulty pipette) were removed. Stimulation index (SI) of cytokine production from both the co-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 $p \leq 0.05$ 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. 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.

2.3 RESULTS

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

IL-8 secretion was used as a metric to distinguish between sensitizers and non-sensitizers. The co-culture of MUTZ-LCs, KCs, and FBs and the MUTZ-LCs alone were both treated with three concentrations per chemical of non-sensitizers and sensitizers (except for SDS, VL, and DNCB, due to cytotoxicity), and supernatants were analyzed by ELISA. The co-culture system demonstrated no significant change in response to non-sensitizers and non-sensitizing irritants. MUTZ-LCs demonstrated a significant decrease in IL-8 secretion in response to non-sensitizing irritant SDS (**Figure 2.2**). The co-culture system accurately identified at least one concentration of every known sensitizer to have significantly increased IL-8 secretion, while the MUTZ-LCs failed to significantly up-regulate IL-8 secretion in response to hapten sensitizer cinnamaldehyde and pre-/pro-haptens cinnamic alcohol, p-phenylenediamine, hydroquinone, and p-benzoquinone (**Figure 2.3**). Across all the concentrations tests of sensitizers and non-sensitizers, the MUTZ-LCs demonstrated 45.1% accuracy in identifying sensitizers by up-regulation of IL-8 secretion and non-sensitizers by insignificant SI relative to vehicle. The co-culture system demonstrated

86.3% accuracy in identifying sensitizers and non-sensitizers by the same metric. To assess accuracy of identification of pre-/pro-haptens, haptens were removed from the count (**Table 2.2**).

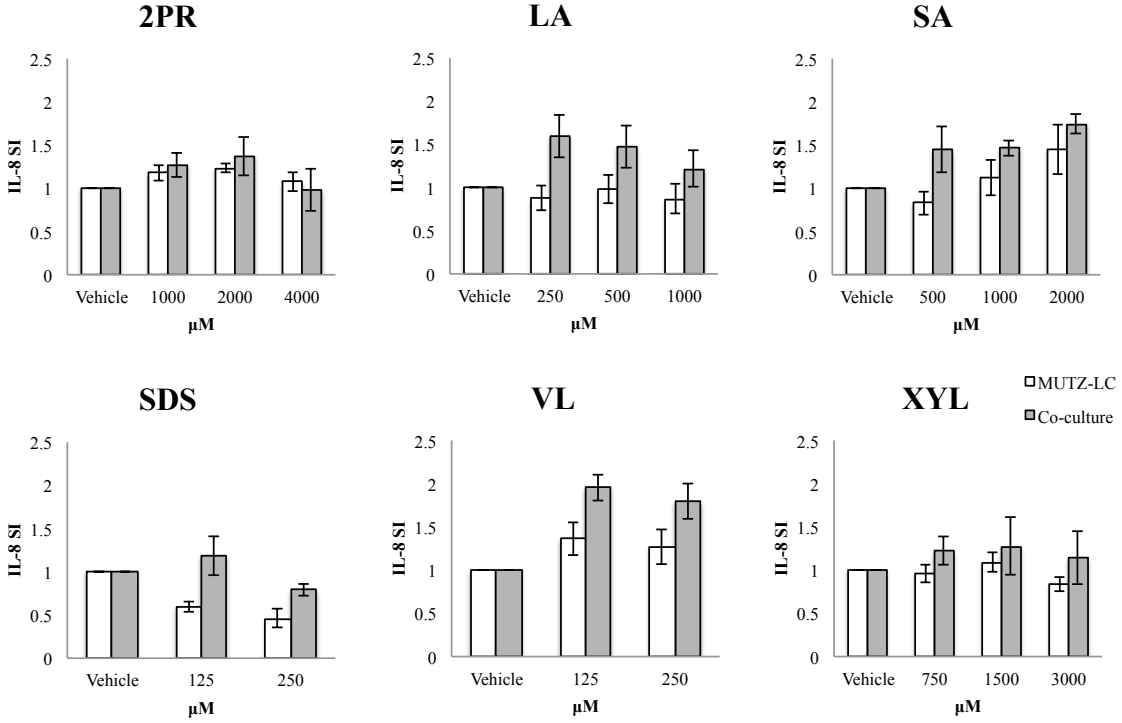

Figure 2.2 IL-8 secretion in response to non-sensitizers. IL-8 secretion by the co-culture and mono-culture MUTZ-LC systems in response to non-sensitizers isopropanol, lactic acid, salicylic acid, sodium dodecyl sulfate, vanillin, and xylene was measured by ELISA. The co-culture demonstrated no significant difference in secretion to any of the six non-sensitizers. MUTZ-LCs alone demonstrated significant decrease in IL-8 secretion in response to non-sensitizing irritant SDS. * indicates p≤.05 by ANOVA, Fisher's LSD post-hoc analysis for n = 4 independent replicates.

Overall Accuracy			
System	**Sensitivity**	**Specificity**	**Accuracy**
MUTZ-LC	25% (9/35)	87.5% (14/16)	45.1% (23/51)
Co-culture	80% (28/35)	100% (16/16)	86.3% (44/51)
Pre-/Pro-Hapten Accuracy			
System	**Sensitivity**	**Specificity**	**Accuracy**
MUTZ-LC	27% (8/30)	87.5% (14/16)	47.8% (22/46)
Co-culture	86.7% (26/30)	100% (16/16)	91.3% (42/46)

Table 2.2 Overall and pre-/pro-hapten accuracy of sensitizer and non-sensitizer identification by IL-8 secretion. 35 concentrations across 12 sensitizers and 16 concentrations across 6 non-sensitizers were assessed for overall analysis. 30 concentrations across 10 pre-/pro-hapten sensitizers were assessed for pre-/pro-hapten accuracy.

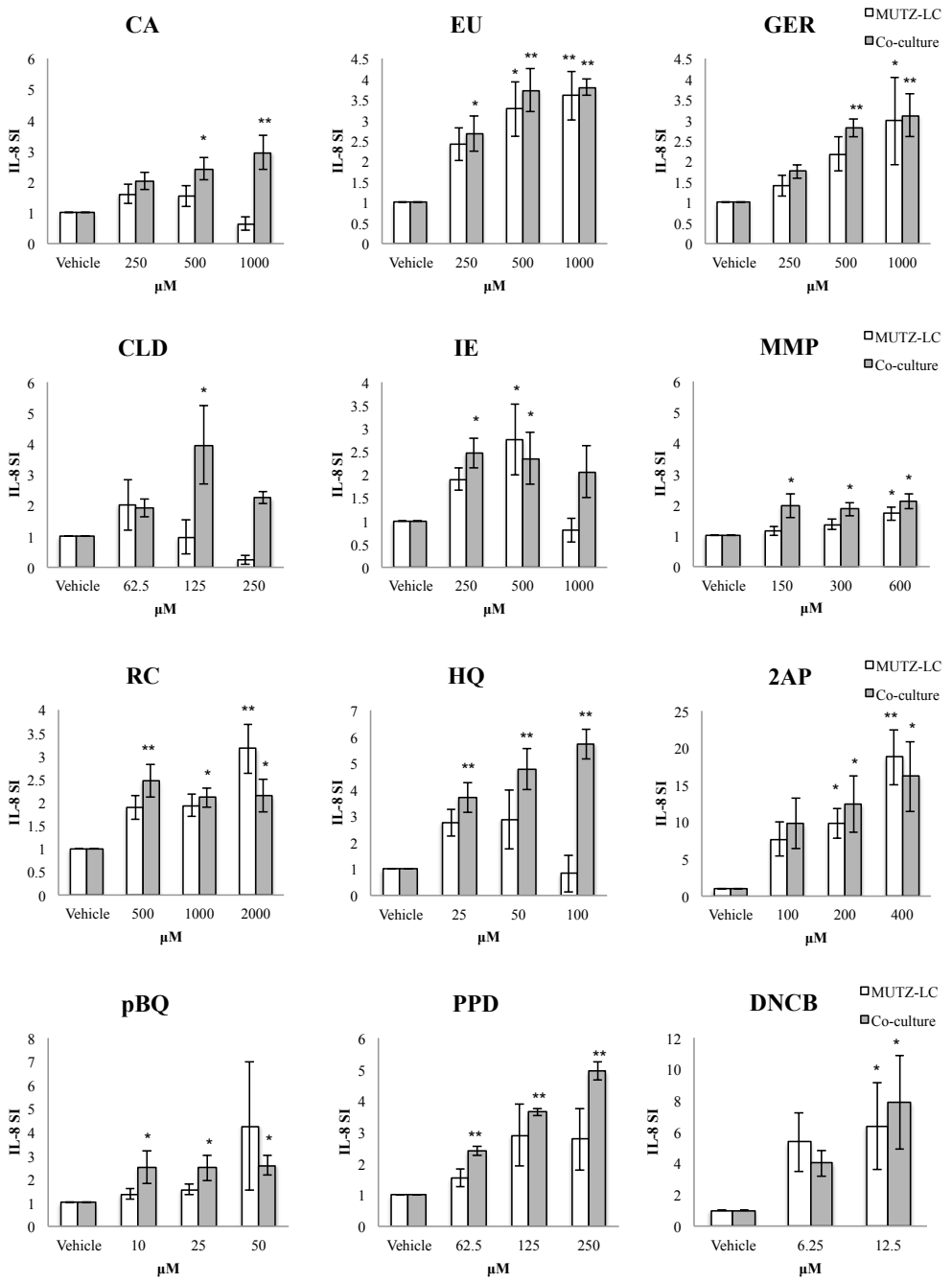

Figure 2.3 IL-8 secretion in response to sensitizers. IL-8 secretion by the co-culture and mono-culture MUTZ-LC systems in response to weak sensitizers cinnamic alcohol, eugenol, and geraniol (first row), moderate sensitizers cinnamaldehyde, isoeugenol, 2-methoxy-4-methylphenol, and resorcinol (second and third rows), strong sensitizers hydroquinone, 2-aminophenol, p-benzoquinone, and p-phenylenediamine (third and fourth rows), and extreme sensitizer 2-4-dinitrochlorobenzene (fourth row) was measured by ELISA. The co-culture demonstrated a significant increase in IL-8 secretion in response to at least 1 concentration of every sensitizer. MUTZ-LCs alone did not up-regulate IL-8 secretion in response to CA, CLD, HQ, pBQ, or PPD. * indicates $p \leq 0.05$ and ** indicates $p \leq 0.005$ by ANOVA, Fisher's LSD post-hoc analysis for n = 4 independent replicates.

ANOVA Analysis of Secretome Data

To identify predictive metrics of sensitization, the 27 biomarker stimulation indices measured in the bio-plex assay were assessed for statistical significance between the two classes "sensitizers" (S) and "non-sensitizers" (NS) by ANOVA, Fisher's LSD post-hoc analysis and ranked by p value. The difference between S and NS was found to be significant in 12 biomarkers measured in the MUTZ-LC mono-culture system, while the difference between S and NS was found to be significant in 20 biomarkers measured in the co-culture system (**Table 2.3**).

MUTZ-LC			**Co-culture**		
Rank	**Metric**	**p value**	**Rank**	**Metric**	**p value**
1	IL-8	<0.0001	1	IL-8	<0.0001
1	IL-1β	<0.0001	1	MIP-1β	<0.0001
1	IL-9	<0.0001	1	GM-CSF	<0.0001
1	IL-17	<0.0001	1	IL-15	<0.0001
1	IL-15	<0.0001	5	MIP-1α	0.0002
6	IL-7	0.0003	6	IL-17	0.0011
7	GM-CSF	0.0004	7	IL-1β	0.0014
8	MCP-1	0.0007	8	G-CSF	0.0048
9	Eotaxin	0.0017	9	RANTES	0.0054
10	IL-2	0.0194	10	VEGF	0.0059
11	TNF-α	0.0196	11	IL-13	0.0085
12	RANTES	0.0257	12	MCP-1	0.0098
			13	IL-10	0.0118
			14	FGF-basic	0.0260
			15	IL-12	0.0268
			16	IFN-γ	0.0286
			17	IL-2	0.0299
			18	IL-9	0.0301
			19	IL-7	0.0339
			20	Eotaxin	0.0430

Table 2.3 Ranking of significant biomarkers for MUTZ-LCs and co-culture by ANOVA p values. By this ranking system, secreted metrics were considered predictive if the secretion difference between all sensitizers and non-sensitizers was significant by ANOVA, Fisher's LSD post-hoc analysis, $p \leq 0.05$.

Support Vector Machine Analysis of Secretome Data

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 use this information to rank each cytokine and select the key features necessary to produce an accurate prediction. As 12 biomarkers were ranked significant

by ANOVA for the MUTZ-LC mono-culture system, we selected the top twelve using the SVM and subsequently smaller combinations and assessed accuracy of prediction. The SVM top twelve secretome biomarkers collected from the mono-culture system in ranked order are 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 (**Table 2.4**). When IL-8 and MIP-1β are used in combination an accuracy of 86.2% is achieved. It should be noted that the SVM identified MIP-1β to be predictive for the MUTZ-LCs while the ANOVA ranking did not ($p = 0.2818$ by ANOVA, Fischer's LSD post-hoc test). Furthermore, attempts to select and rank more than twelve biomarkers for the MUTZ-LC system using the SVM failed; despite increasing the maximum-allowed iterations 5-fold, the data could not converge to maximize margin distance and minimize error.

Rank	Metric	Margin Distance
1	IL-8	0.14418
2	MIP-1β	0.13576
3	IL-9	0.12928
4	IL-17	0.12578
5	MIP-1α	0.12300
6	IL-1β	0.12228
7	IL-15	0.12175
8	RANTES	0.11879
9	GM-CSF	0.11867
10	MCP-1	0.11678
11	IL-7	0.11158
12	Eotaxin	0.10952

Table 2.4 Top ranking biomarkers of prediction by SVM-computed margin distances for MUTZ-LC mono-culture.

Though 20 biomarkers were ranked as significant by ANOVA for the co-culture system, the SVM did not succeed in maximizing margin distances and minimizing error, and thus could not compute and rank the top 20 biomarkers from the co-culture system, despite increasing the maximum-allowed iterations 500-fold. The top twelve secretome biomarkers collected from the co-culture system in ranked order are IL-8, MIP-1β, GM-CSF, RANTES, IL-15, MCP-1, MIP-

1α, IL-17, VEGF, IL-1β, G-CSF, and IL-13, achieving 86.7% accuracy in combination (**Table 2.5**). The top three cytokines, IL-8, MIP-1β, and GM-CSF, offer a classification model with 91.1% accuracy when used in combination (**Table 2.6**).

Rank	Metric	Margin Distance
1	IL-8	0.17234
2	MIP-1β	0.13126
3	GM-CSF	0.13101
4	RANTES	0.12753
5	IL-15	0.12057
6	MCP-1	0.11953
7	MIP-1α	0.11682
8	IL-17	0.11423
9	VEGF	0.11337
10	IL-1β	0.11279
11	G-CSF	0.11116
12	IL-13	0.11110

Table 2.5 Top ranking biomarkers of prediction by SVM-computed margin distance for the co-culture system.

MUTZ-LCs alone		**Co-culture**	
Metric	*Accuracy*	*Metric*	*Accuracy*
Top 12	**87.2%**	Top 12	86.7%
Top 10	86.2%	Top 10	86.7 %
Top 5	83.5%	Top 4	90%
Top 3	83%	Top 3	**91.1%**
Top 2	86.2%	Top 2	87.8%
IL-8	84%	IL-8	86.7%

Table 2.6 SVM accuracy for both *in vitro* systems using feature selection. The highest achievable accuracy for each system is noted in bold print. The top 12 biomarkers from the MUTZ-LC secretome achieve the highest accuracy in combination while the top 3 biomarkers from the co-culture secretome achieve the highest accuracy in combination. For both systems, a combination of biomarkers achieves higher accuracy than the top ranked biomarker (IL-8) alone.

The SVM also calculated the sensitivity and specificity of each classification model (**Table 2.7**).

System	Sensitivity	Specificity	Correct Rate
MUTZ-LCs alone	95.5%	79.6%	87.2%
Co-culture	92.7%	89.8%	91.1%

Table 2.7 Overall sensitivity, specificity, and correct rate (accuracy) predicted by the SVM. Calculated for the top 12 biomarkers from the MUTZ-LC mono-culture system and the top 3 biomarkers from the co-culture system.

The SVM was further 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

co-culture systems. IL-8 and MIP-1β achieve 87.2% accuracy in combination for the MUTZ-LC mono-culture system, while IL-8, RANTES, and GM-CSF achieve 90.2% accuracy for the co-culture system (**Table 2.8**).

System	Sensitivity	Specificity	Correct Rate
MUTZ-LC alone	93.3%	80.5%	87.2%
Co-culture	92.7%	87.8%	90.2%

Table 2.8 Pre-/pro-hapten sensitivity, specificity, and correct rate predicted by the SVM. Calculated for the top 2 biomarkers from the MUTZ-LC mono-culture system and the top 3 biomarkers from the co-culture system as predictive of pre-/pro-hapten sensitization.

Classification Trees for Stimulation Index Threshold Identification

Biomarker SI after 48 hours of chemical treatment was calculated and four data points from every chemical concentration were averaged and used in the classification tree algorithm written in R. 18 averaged non-sensitizer data points and 35 averaged sensitizer data points were assessed using our classification algorithm. The classification tree for IL-8 SI in MUTZ-LCs alone identified a threshold of 1.366 with a p value of 0.011 (**Figure 2.4a**). Based on this threshold, the MUTZ-LCs alone misclassify CA at 1000 μM, CLD at 125 and 250 μM, IE at 1000μM, MMP at 150 and 300 μM, and HQ at 100 μM as non-sensitizers and SA at 2000 μM as a sensitizer, resulting in an overall accuracy of 86.8% (**Table 2.9**). The classification tree for IL-8 SI in co-culture identified a threshold of 1.743 with a p value of 0.002 (**Figure 2.4b**). Based on this threshold, the co-culture system misclassifies SA at 2000 μM and VL as sensitizers, resulting in an overall accuracy of 94.3% (**Table 2.9**). Based on the computed threshold, the co-culture system did not predict any false negatives.

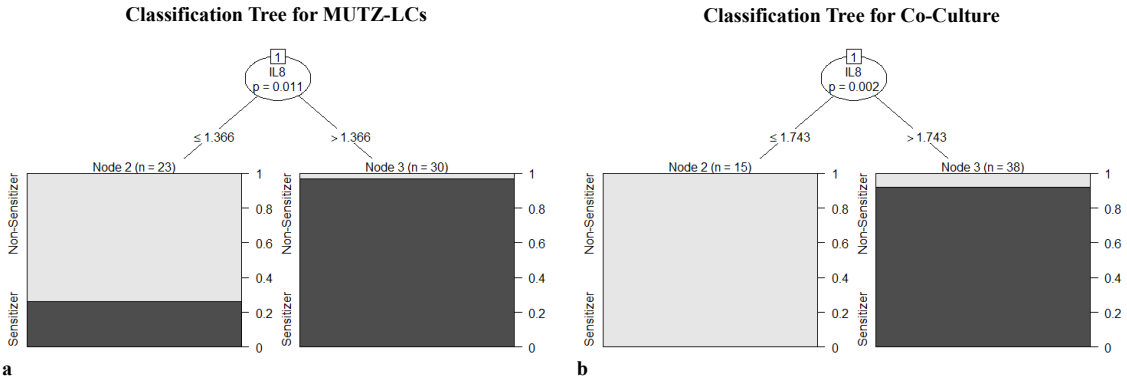

Figure 2.4 IL-8 classification trees generated in R for MUTZ-LCs and co-culture. (a) A classification tree for MUTZ-LCs alone identified a threshold of 1.366 SI to distinguish between sensitizer and non-sensitizer using IL-8 as a biomarker. (b) A classification tree for the co-culture system identified a threshold of 1.743 SI to distinguish between sensitizer and non-sensitizer using IL-8 as a biomarker.

System	Biomarker	Sensitivity	Specificity	Accuracy
MUTZ-LC	IL-8	96.7%	73.9%	86.8%
Co-culture	IL-8	92.1%	100%	94.3%

Table 2.9 Accuracy of sensitizer prediction based on threshold values from classification trees.

IL-8 Secretion to Evaluate Sensitizer Potency

We tested 3 weak, 4 moderate, 3 strong, and 2 extreme sensitizers (based on LLNA potency classification) with our MUTZ-LC and co-culture systems. We grouped the IL-8 SI data from the weak and moderate sensitizers together (W/M) and the data from the strong and extreme sensitizers together (S/E) and used ANOVA and Fisher's LSD post-hoc analysis to determine statistical significance ($p \leq 0.05$) between responses to non-sensitizers, W/M sensitizers, and S/E sensitizers. For both systems, both grouped potency categories were significantly up-regulated relative to non-sensitizers, and S/E were significantly increased relative to W/M.

Using the same groupings (NS, W/M, S/E), 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 our 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 (**Figure 2.5**). Based on these thresholds, the MUTZ-LCs alone misclassify CA at 1000 µM, CLD at 125 and 250 µM, IE at 1000µM, 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 2.10**).

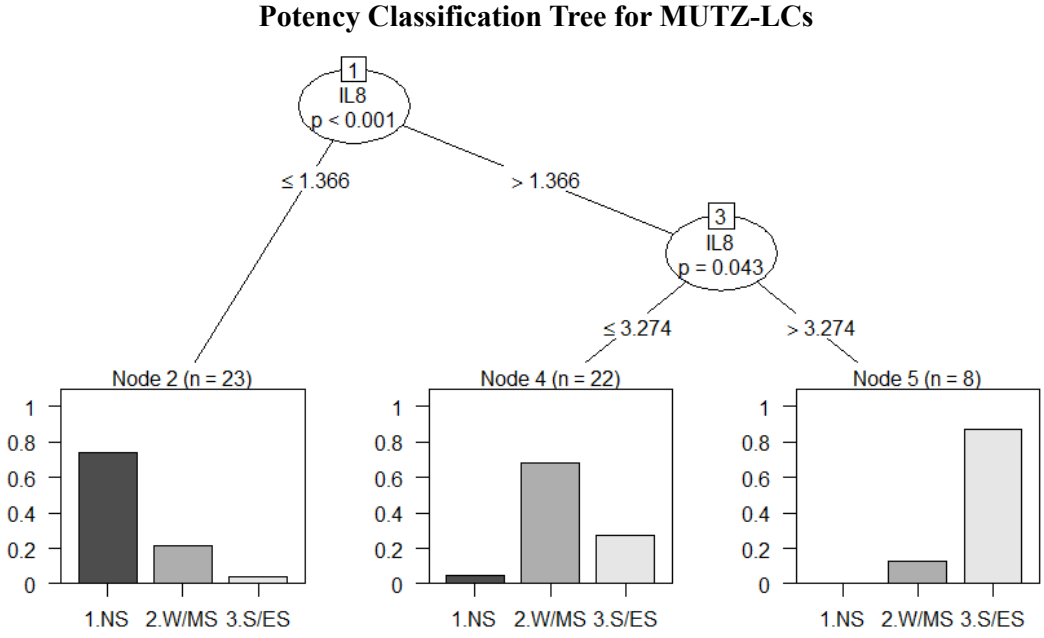

Figure 2.5 IL-8 potency classification tree generated in R for MUTZ-LCs. The algorithm identified a threshold of 1.366 SI to distinguish between non-sensitizers and all sensitizers and a break point within the sensitizer group of 3.274 SI to distinguish between W/M and S/E sensitizers.

The classification tree for IL-8 SI in co-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 (**Figure 2.6**). Based on these thresholds, the co-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% (**Table 2.10**). Based on the computed threshold, the co-culture system did not predict any false negatives.

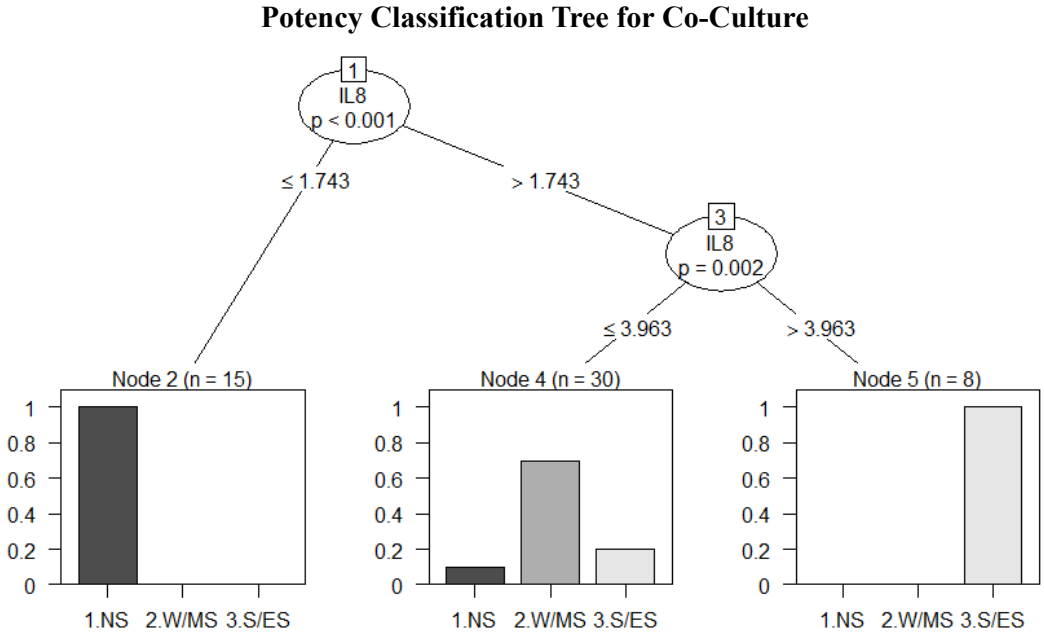

Figure 2.6 IL-8 potency classification tree generated in R for co-culture. The algorithm identified a threshold of 1.743 SI to distinguish between non-sensitizers and all sensitizers and a break point within the sensitizer group of 3.963 SI to distinguish between W/M and S/E sensitizers.

	Non-Sensitizers	W/M Sensitizers	S/E Sensitizers
MUTZ-LCs			
Sensitivity	73.9%	68.2%	87.5%
Specificity	96.7%	80.7%	84.4%
Accuracy	85.3%	74.4%	86%
Co-culture			
Sensitivity	100%	70%	100%
Specificity	92.1%	100%	86.7%
Accuracy	96.1%	85%	93.3%

Table 2.10 Accuracy of potency prediction based on threshold values from classification trees.

2.4 DISCUSSION

There is an established need for predictive sensitizer screening assays. Currently, the *in vitro* assays validated or undergoing validation by the OECD have particularly low accuracy in predicting pre- and pro-hapten sensitizers, which require chemical transformation or metabolic activation prior to inducing sensitization. Full-thickness skin equivalents have shown promise in addressing this limitation, but are expensive and low-throughput. Therefore, we attempted to mimic the environment of a 3D full-thickness skin model in a 2D co-culture system that could be implemented in 96-well plates. This co-culture system was compared to a mono-culture system to demonstrate the benefits of incorporating skin cells.

IL-8 secretion was chosen as a starting metric of sensitization based on literature precedence[11-14,17]. Neither the MUTZ-LCs alone nor the co-culture system demonstrated statistically significant up-regulation of IL-8 secretion in response to non-sensitizers or non-sensitizing irritants. Interestingly, MUTZ-LCs alone were found to down-regulate IL-8 significantly relative to vehicle in response to the surfactant SDS. This phenomenon was also seen with moDCs *in vitro*[17]. This is in contrast to KeratinoSens™ and U-SENS™, which identified SDS as a false positive[18-19]. SDS was also identified as a false positive in the LLNA[6]. In the optimization of the IL-8 Luc assay, detergents (including SDS, Tween-80, and benzalkonium chloride) were included in an exclusion criterion, as detergents were found to test positive in that assay[13]. It is possible that MUTZ present a more physiologically relevant DC-like cell than the U-937 and THP-1 lines and therefore do not predict SDS to be a sensitizer.

By traditional ANOVA analysis, MUTZ-LCs failed to significantly up-regulate IL-8 secretion in response to a number of sensitizers, including 1 hapten (CLD) and 4 pre-/pro-haptens (CA, HQ, pBQ, and PPD). The co-culture system demonstrated significant up-regulation of IL-8 secretion in response to at least one concentration of every chemical. We believe the reason for this discrepancy is the lack of metabolic components necessary to convert pre- and pro-haptens to active haptens in the MUTZ mono-culture, as it has been shown that DCs express lower CYP

levels compared to normal human skin; the metabolic enzymes are produced primarily by keratinocytes[20]. The co-culture did not significantly up-regulate IL-8 in response to the lowest concentrations of some of the weak sensitizers. It is possible that at those low concentrations the chemicals are safe and would not induce sensitization. Such conclusions are critical for the cosmetics industry, as many aromatic chemicals are pre- or pro-hapten sensitizers, and a maximum concentration permitted in cosmetic formulations must be codified for safety. The co-culture also did not significantly up-regulate IL-8 secretion in response to the highest concentration of isoeugenol. We suspect this may be an effect of high cytotoxicity, as it was later found that the highest concentration used was more cytotoxic than originally thought (i.e. <50% viable, unpublished observations).

The initial IL-8 secretion results demonstrated that the co-culture system is superior to MUTZ-LCs alone for sensitizer identification, and pre- and pro-hapten sensitizers in particular. At 86% accuracy, the system left room for improvement, and so a 27-cytokine multi-plex assay was used to analyze the secretomes of the co-culture system and the MUTZ mono-culture to identify additional predictive metrics that could improve accuracy in combination. Preliminary analysis of the multi-plex data by ANOVA found 12/27 biomarkers demonstrated significant difference between the positive (sensitizer) class and negative (non-sensitizer) class in the MUTZ-LC system and 20/27 in the co-culture system. This analysis did not offer a means to calculate prediction accuracy, but would offer a point of comparison to the output from the SVM. In addition, it would be extremely inefficient and expensive to analyze 12 or 20 biomarkers by ELISA, and so ANOVA identification of significant biomarkers did not offer improvement to the assay. 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 combinations were initially chosen based on the number of biomarkers identified by ANOVA to be significant; however, the SVM was unable to maximize the margin distances for 20 biomarkers in combination from the

co-culture system. This seems to indicate that the algorithm is unable to converge to a maximum margin distance for each of the 20 biomarkers, i.e. such a predictive classification model does not exist. The SVM was able to build a classification model for the MUTZ-LCs with the top 12 biomarkers; indeed, this model had the highest accuracy of 87%. The top 2 biomarkers (IL-8 and MIP-1β), though, achieved an accuracy of 86% in combination, and this model is more efficient and less expensive than one that requires testing 12 biomarkers for little trade-off in accuracy. It is interesting to note that the top 12 biomarkers identified by ANOVA do not match the top 12 identified by the SVM. Specifically, MIP-1β and MIP-1α were not found to be significant by ANOVA. This is due to large variance in the raw data and identifies a potential pitfall in the SVM: variance within a class will not disturb the algorithmic efforts to maximize margin distance so long as that variance does not overlap into the other class. The co-culture system achieved its highest accuracy 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; we have already established its utility and precedence as a predictive biomarker of sensitization. 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; in that study, prediction accuracy was increased by combining measurement of MIP-1β secretion and CD86 expression[21]. 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 co-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 is also shown that the co-culture system outperforms the 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 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[22]. 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 co-culture system achieves similar accuracy (91%) in prediction of skin sensitizers to the full-thickness skin model (92%)[15]. This suggests that the co-culture system could offer a suitable and more cost-effective alternative to full-thickness skin models without sacrificing accuracy. It is interesting to note that the predictive metrics identified by the SVM for the full-thickness skin model differ from those for the co-culture system. It is possible that primary keratinocytes will behave slightly differently from a keratinocyte cell line, but this warrants further study.

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 sensitizer and non-sensitizer. Using these thresholds, it was easy to identify the misclassified chemicals. The MUTZ-LCs alone identified a number of false negatives. However, the c-tree model appears to be more sensitive than ANOVA analysis of IL-8 secretion, as the number of false negatives identified by the c-tree is much less than by ANOVA (7 and 26, respectively). This is even more apparent in the c-tree for the co-culture system, where no false negatives were identified. The chemicals identified as false positives by the c-tree for the co-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[19,23-24]. 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. It is

unclear to us why the highest concentration of salicylic acid was classified as a sensitizer for both the co-culture and mono-culture systems. It is possible that the high concentration induced cellular stress that triggered a response; this will require further study. Overall, the use of classification trees introduced a more sensitive statistical approach to classifying chemicals based on IL-8 stimulation index, and therefore it was pursued for potency analysis.

In the potency prediction model, a 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)[25]. The thresholds identified between sensitizers and non-sensitizers did not change in the potency tree, rather, another node was added to the sensitizer branch. The contributing factors to the potency misclassification of varying concentrations of pBQ and PPD for both are not entirely clear. It is possible that combining prediction with MIP-1β and GM-CSF will improve potency prediction. It is worth mentioning that the co-culture system correctly classified all W/M sensitizers. 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[26]. Overall, the c-tree method demonstrates promise for potency prediction using cytokine SI, but requires further study.

We developed a co-culture system that can predict pre- and pro-haptens with higher accuracy than the currently validated and pre-validated screening assays and identified a molecular signature using a support vector machine that offers higher accuracy than one biomarker alone. Furthermore, we demonstrated that IL-8 secretion may be used to predict sensitizer potency. However, this assay is not without limitations. The original incubation times were optimized for the full-thickness skin model. For consistency, 48 hours was maintained to be the incubation time for chemical treatment. However, this time point may not be best for a 2D culture model. A chemical may need more time to penetrate and activate LCs in a full-thickness model, but in a 2D

model the chemicals are added directly to the culture, introducing immediate attack on the cells. Another concern is the reliability of Alamar Blue for viability assessment. The reduction of Alamar Blue is enzymatic in nature. As we are treating our cells with sensitizing agents, it is likely that expression of metabolic enzymes is induced in treated conditions compared to vehicle, which could impact the read-out. Furthermore, the same time-point was used for both the co-culture system and the MUTZ-LC mono-culture. Every cell line has unique metabolic properties and must be individually characterized to determine the experimental parameters, such as incubation time and dilution factor, for optimal conversion of Alamar Blue[27]. The Alamar Blue data was used to calculate the cytokine stimulation index (SI), so the SI reflected the cytokine secretion per viable cell relative to vehicle. Therefore, a significant increase in SI indicates cellular up-regulation of that cytokine. Unreliable viability data could raise questions about our conclusions. Finally, identification of other predictive metrics outside the realm of cytokine secretion could be useful for mechanistic studies of ACD and further reduce the use of animals in scientific testing by avoiding animal-derived monoclonal antibodies employed in ELISAs. It should be noted that it is widely established that a single assay is insufficient to screen all skin sensitizers. A particular limitation of our co-culture system that cannot be overcome is the testing of chemicals insoluble in water, as all 2D culture platforms require water-based culture medium. This limitation is best addressed using epidermal equivalents or full-thickness skin models, which could offer prediction of chemicals that are insoluble or unstable in culture medium[28]. By addressing these limitations within reason and expanding the predictive capability of our assay beyond secreted metrics, we can introduce a high-accuracy, cost-effective, high-throughput, multi-metric assay for screening skin sensitizers.

2.5 REFERENCES

1. Thyssen, J. P., et al. (2007). "The epidemiology of contact allergy in the general population--prevalence and main findings." Contact Dermatitis **57**(5): 287-299.

2. Enk, A. H. (1997). "Allergic contact dermatitis: understanding the immune response and potential for targeted therapy using cytokines." Mol Med Today **3**(10): 423-428.
3. Divkovic, M., et al. (2005). "Hapten-protein binding: from theory to practical application in the in vitro prediction of skin sensitization." Contact Dermatitis **53**(4): 189-200.
4. Peiser, M., et al. (2012). "Allergic contact dermatitis: epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. Current knowledge assembled at an international workshop at BfR, Germany." Cell Mol Life Sci **69**(5): 763-781.
5. Karlberg, A. T., et al. (2008). "Allergic contact dermatitis--formation, structural requirements, and reactivity of skin sensitizers." Chem Res Toxicol **21**(1): 53-69.
6. Anderson, S. E., et al. (2011). "The LLNA: A Brief Review of Recent Advances and Limitations." J Allergy (Cairo) **2011**: 424203.
7. Sharma, N. S., et al. (2012). "Perspectives on Non-Animal Alternatives for Assessing Sensitization Potential in Allergic Contact Dermatitis." Cell Mol Bioeng **5**(1): 52-72.
8. Larsen, C. G., et al. (1989). "Production of interleukin-8 by human dermal fibroblasts and keratinocytes in response to interleukin-1 or tumour necrosis factor." Immunology **68**(1): 31-36.
9. Taub, D. D., et al. (1996). "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.
10. Larsen, C. G., et al. (1995). "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.
11. Bergstrom, M. A., et al. (2007). "A skin-like cytochrome P450 cocktail activates prohaptens to contact allergenic metabolites." J Invest Dermatol **127**(5): 1145-1153.
12. Nelissen, I., et al. (2009). "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.
13. Kimura, Y., et al. (2015). "Optimization of the IL-8 Luc assay as an in vitro test for skin sensitization." Toxicol In Vitro **29**(7): 1816-1830.
14. Ouwehand, K., et al. (2010). "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.
15. Lee, S., et al. (2014). "Predicting full thickness skin sensitization using a support vector machine." Toxicol In Vitro **28**(8): 1413-1423.
16. Asturiol, D., et al. (2016). "Consensus of classification trees for skin sensitisation hazard prediction." Toxicol In Vitro **36**: 197-209.
17. Toebak, M. J., et al. (2006). "CXCL8 secretion by dendritic cells predicts contact allergens from irritants." Toxicol In Vitro **20**(1): 117-124.
18. Natsch, A., et al. (2011). "The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict sensitizers in vitro: Results of a ring-study in five laboratories." Toxicol In Vitro. **25**(3): 733-744.
19. Piroird, C., et al. (2015). "The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization." Toxicol In Vitro **29**(5): 901-916.
20. Chipinda, I., et al. (2011). "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.
21. Hirota, M. and O. Moro (2006). "MIP-1beta, a novel biomarker for in vitro sensitization test using human monocytic cell line." Toxicol In Vitro **20**(5): 736-742.
22. Corsini, E. and C. L. Galli (2000). "Epidermal cytokines in experimental contact dermatitis." Toxicology **142**(3): 203-211.
23. Saito, K., et al. (2013). "Development of an in vitro skin sensitization test based on ROS production in THP-1 cells." Toxicol In Vitro **27**(2): 857-863.
24. Takenouchi, O., et al. (2015). "Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals." J Appl Toxicol **35**(11): 1318-1332.

25. Committee on the Design and Evaluation of Safer Chemical Substitutions: A Framework 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 Oct 29. Appendix D, Overview of the GHS Classification Scheme in Hazard Classification. Available from: https://www.ncbi.nlm.nih.gov/books/NBK253967/
26. Basketter, D. A. and I. Kimber (2010). "Skin sensitization, false positives and false negatives: experience with guinea pig assays." J Appl Toxicol **30**(5): 381-386.
27. Nakayama, G. R., et al. (1997). "Assessment of the Alamar Blue assay for cellular growth and viability in vitro." J Immunol Methods **204**(2): 205-208.
28. Gibbs, S., et al. (2013). "An epidermal equivalent assay for identification and ranking potency of contact sensitizers." Toxicol Appl Pharmacol **272**(2): 529-541.

CHAPTER 3: EVALUATING THE INDIVIDUAL CELLULAR CONTRIBUTION OF CXCL8 IN RESPONSE TO SKIN SENSITIZERS

3.1 INTRODUCTION

Allergic contact dermatitis (ACD) is an inflammatory skin disease mediated by T-cells. The prevalence of contact sensitizers necessitates screening assays. The limitations and ban of animal testing has motivated research in alternative methods, but currently validated and pre-validated alternative screening assays are particularly limited in their prediction of pre- and pro-haptens sensitizers, which require chemical conversion or metabolic activation prior to inducing sensitization. To address these limitations we previously developed a co-culture method using MUTZ-3-derived Langerhans cells (MUTZ-LCs), HaCaT keratinocytes (KCs), and dermal fibroblasts (FBs). In chapter 2 this method was expanded and paired with a support vector machine to build a classification model for more accurate prediction via a multi-metric signature. The method was further assessed for accuracy in potency prediction.

Though this co-culture system presents a promising *in vitro* assay for skin sensitizer prediction, it could be further expanded to include multiple measurable metrics across different steps of the adverse outcome pathway (AOP) and sub-stages within those steps. Real-time analysis could offer opportunity to assess cellular changes in response to sensitizers as they occur. However, to identify additional metrics, it is helpful to first elucidate the contributions and roles of each cell type in the inflammatory response to skin sensitizers. For example, prior to generating an IL-8 reporter in a dendritic cell line for a co-culture system, it would be prudent to establish evidence that the dendritic cells are the primary producers of IL-8. Keratinocytes have been shown to secrete pro-inflammatory cytokines and play diverse roles in dermatotoxic actions[1]. In particular, KCs have demonstrated increases in synthesis and secretion of IL-1, TNF-α, and IL-8 in response to non-sensitizing irritants[2]. Though we have established that IL-8 is up-regulated in response to skin sensitizers in our co-culture system, it is not clear which cell(s) are producing the IL-8.

Understanding which cells are most involved in the pro-inflammatory response will offer a starting point to investigate additional pro-inflammatory metrics of sensitization and could enhance our understanding of cross-talk between cells in the sensitization process.

3.2 MATERIALS AND METHODS

Cell Culture

HaCaT keratinocytes, human dermal fibroblasts, and MUTZ-3 cells were cultured as described in chapter 2.

Chemicals and Reagents

Test chemicals included weak, pre-/pro-hapten eugenol (EU), extreme hapten 2-4-dinitrochlorobenzene (DNCB), non-sensitizer xylene (XYL), non-sensitizing irritant sodium dodecyl sulfate (SDS), and the vehicle (0.1% DMSO) (**Table 3.1**). Isoeugenol (IE) and hydroquinone (HQ) had been included, but the chosen concentrations were later determined to be too cytotoxic (<50% viable), and thus these chemicals were eliminated from the test panel and results.

Chemical	Abbreviation	Class	Potency	Concentration (µM)
2-4-dinitrochlorobenzene	DNCB	Hapten	Extreme	12.5
Eugenol	EU	Pre-/Pro-Hapten	Weak	1000
Xylene	XYL	Non-Sensitizer	-	3000
Sodium Dodecyl Sulfate	SDS	Non-Sensitizer	Irritant	500

Table 3.1 Panel of chemicals evaluated.

Co-cultures and Mono-cultures of HaCaT Keratinocytes, Dermal Fibroblasts, and MUTZ-3 Langerhans Cells

HaCaT KCs and FBs were plated in mono-culture and in co-culture in 96-well plates in complete DMEM the night before the start of the experiment (i.e. day 6 of MUTZ-LC differentiation) to allow ample time for the cells to adhere. On day 7 the wells were washed with maturation medium and fully differentiated MUTZ-LCs were plated in mono-culture and in every co-culture

combination possible, such that the total number of cells in every well was $2.5x10^4$ in a volume of 220 µL of chemical treatment or vehicle, and 7 different culture conditions were treated: MUTZ-LCs alone, KCs alone, FBs alone, KCs+FBs, MUTZ-LCs+KCs, MUTZ-LCs+FBs, and MUTZ-LCs+KCs+FBs (referred to as tri-culture in this chapter). Wells containing both skin cells and MUTZ-LCs contained a 1:1 ratio of (all) skin cells to immune cells (**Figure 3.1**). Duplicate plates of all cultures were prepared so one plate was utilized to assess viability and the other was used to collect supernatant for IL-8 secretion analysis. Cells incubated for 48 hours at 37°C and 5% CO_2 and then viability was assessed using Alamar Blue analysis and supernatant was collected and stored at -20°C for future ELISA analysis.

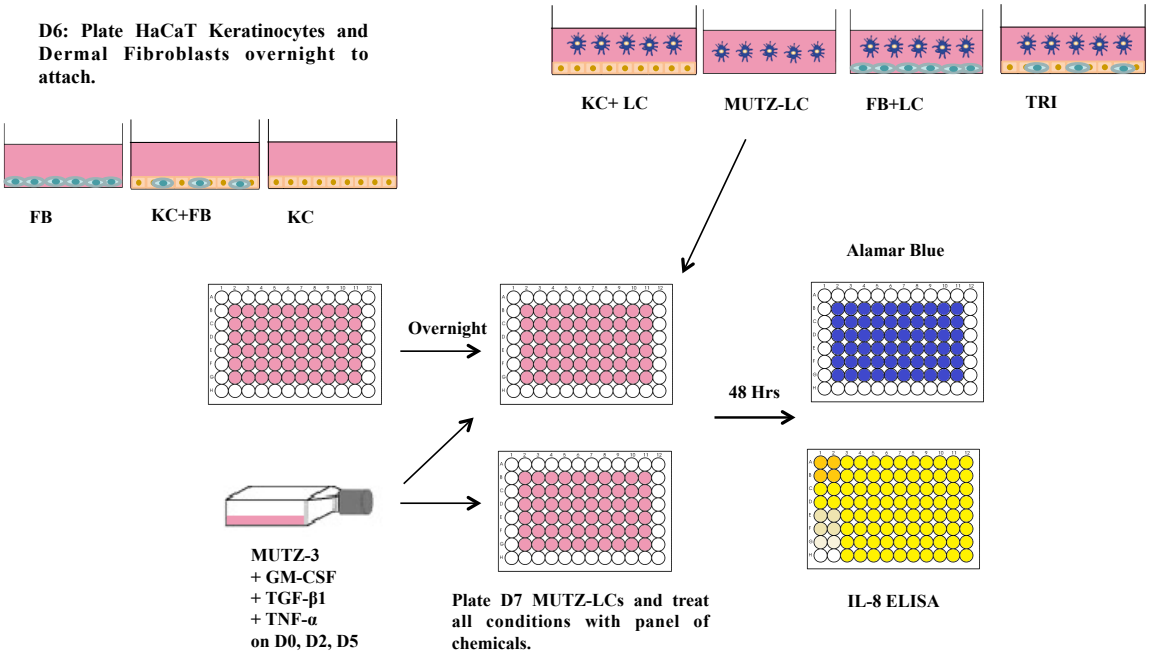

Figure 3.1 Experimental schematic for co-cultures and mono-cultures of MUTZ-LCs, HaCaT KCs, and dermal FBs for IL-8 contribution studies.

Viability

Viability was assessed and quantified as described in chapter 2.

IL-8 Secretion

Supernatants collected from all co-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.

Data Acquisition and Analysis

A total of four independent experiments were conducted where each condition (except the tri-culture) was tested in triplicate per experiment and used for subsequent data analysis. Raw secretion (measured in pg/mL) and viability were averaged across triplicates. Stimulation index (SI) of cytokine production from both the co-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 raw IL-8 secretion and IL-8 SI measured by ELISA was determined at $p \leq 0.05$ using ANOVA and Fisher's least significant difference (LSD) post-hoc analysis in Kaleidagraph.

3.3 RESULTS

Evaluation of IL-8 Secretion in Mono-Cultures and Co-cultures of KCs, FBs, and MUTZ-LCs

IL-8 secretion in response to sensitizers and non-sensitizers was used as a metric to elucidate the necessary and active components for accurate prediction of skin sensitization. Mono-cultures of skin cells demonstrated low levels of IL-8 secretion across all conditions and no significant up-regulation of IL-8 secretion in response to skin sensitizers. KCs alone showed significant decrease in IL-8 secretion in response to EU and SDS. Co-culture of KCs and FBs also demonstrated low levels of IL-8 secretion across all conditions and no significant up-regulation of

IL-8 secretion in response to skin sensitizers, but a significant decrease in IL-8 secretion in response to EU (**Figure 3.2**).

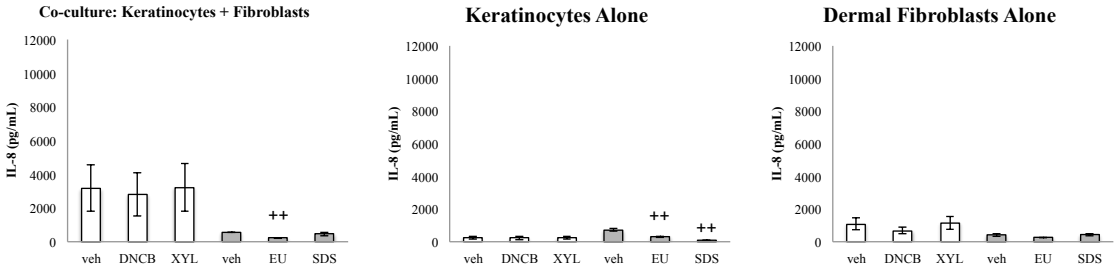

Figure 3.2 Raw IL-8 secretion in mono-cultures and co-culture of skin cells in response to sensitizers and non-sensitizers. ++ represents a significant decrease in IL-8 secretion with p value p≤0.005 calculated by ANOVA, Fisher's LSD post-hoc analysis for n = 3 independent replicates.

Mono-culture of MUTZ-LCs and co-culture of MUTZ-LCs with skin cells demonstrated up-regulation of IL-8 secretion in response to DNCB. Only MUTZ-LCs alone demonstrated significant response to EU in raw IL-8 secretion (**Figure 3.3**).

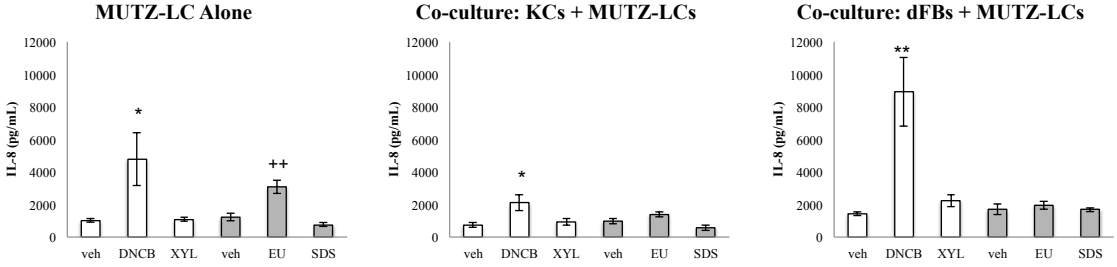

Figure 3.3 Raw IL-8 secretion in MUTZ-LC mono-culture and co-culture of MUTZ-LCs and skin cells in response to sensitizers and non-sensitizers. * indicates p≤0.05 and ** indicates p≤0.005 for n = 4 independent replicates. ++ indicates a significant increase in IL-8 secretion with p value p≤0.005 for n = 3 independent replicates. All statistics calculated by ANOVA, Fisher's LSD post-hoc analysis.

IL-8 SI was calculated to determine up-regulation of IL-8 relative to vehicle condition. IL-8 SI was significantly increased relative to vehicle in response to sensitizers only in the presence of MUTZ-LCs (in every combination). IL-8 was not up-regulated in response to non-sensitizers in the presence of MUTZ-LCs. IL-8 SI in response to xylene was increased in co-culture of KCs and FBs. Interestingly, IL-8 SI was significantly decreased in response to SDS in the presence of KCs (**Figure 3.4**).

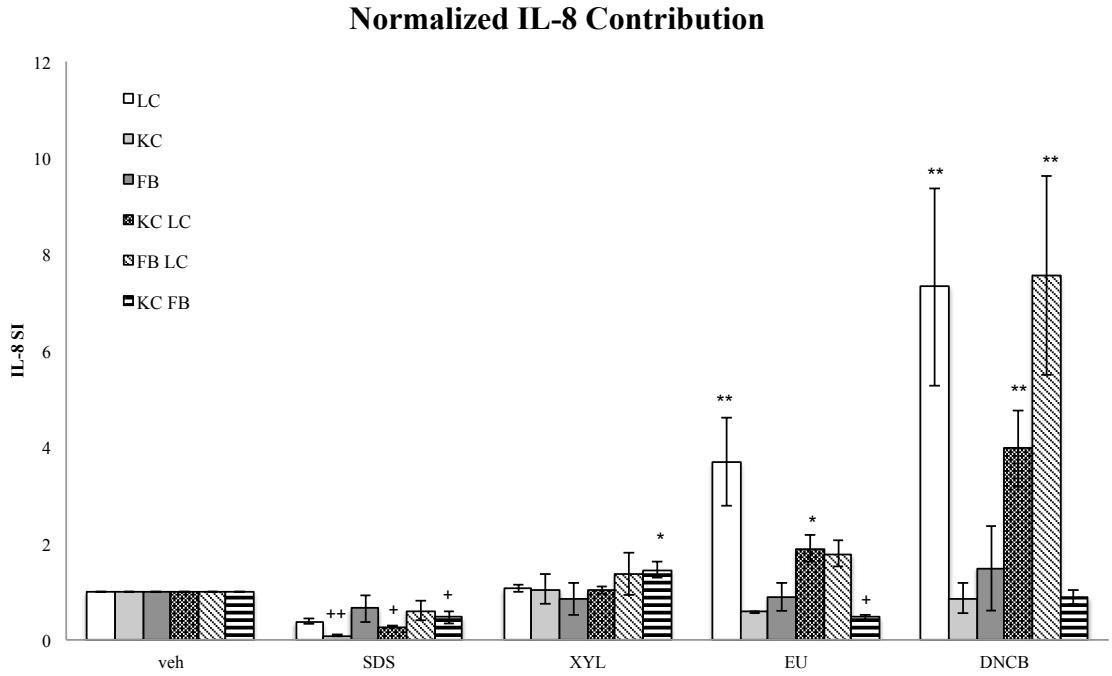

Figure 3.4 IL-8 SI in mono-cultures and co-cultures of MUTZ-LCs, KCs, and FBs in response to sensitizers and non-sensitizers. * indicates p≤0.05 and ** indicates p≤0.005 for increase in IL-8 SI for n = 3-4 independent replicates. + indicates p≤0.05 and ++ indicates p≤0.005 for decrease in IL-8 SI for n = 3 independent replicates.

3.4 DISCUSSION

We previously developed an in vitro co-culture system that mimics the components of a full-thickness skin model in a 2D platform to screen skin sensitizers. Chapter 2 describes the identification and assessment of predictive biomarkers for an established assay using this co-culture system. An inflammatory molecular signature from secretome offers higher prediction accuracy than one inflammatory biomarker alone. The purpose of these studies was to identify the requisite or active cell(s) in the pro-inflammatory stage of skin sensitization in our culture system. This information can be used for informed targeting to identify additional markers of the inflammatory response that could be predictive of sensitization.

IL-8 secretion analysis of the skin cells in our co-culture system in every possible mono-culture and co-culture combination demonstrated that skin cells secrete very little IL-8 in general, and do not up-regulate this secretion in response to sensitizers or non-sensitizers. In fact, a decrease in

secretion was noted in some conditions. This could reflect lower viability in chemically-treated conditions or an actually down-regulation of IL-8 expression and secretion in response to non-sensitizing irritants. A similar trend has been noted in moDCs, but not skin cells[3]. Raw secretion of IL-8 in MUTZ-LCs in mono-culture and co-culture with skin cells was low in non-sensitizer treated and control conditions. DNCB-treated conditions containing MUTZ-LCs demonstrated a significant increase in IL-8 secretion to thousands of pg/mL. Interestingly, co-culture of KCs+MUTZ-LCs and FBs+MUTZ-LCs did not result in a statistically significant increase in raw secretion in response to eugenol. This may reflect variance between experiments due to biological conditions that would increase the standard deviation and thus reduce statistical significance between a chemical condition and the vehicle. It should also be noted that the number of MUTZ cells in both these co-culture conditions was half the number in the mono-culture. If MUTZ cells are the primary producers of IL-8, fewer MUTZ could result in less up-regulation of IL-8. To reduce variance and further take into account each condition's viability (i.e. the quantity of live, IL-8-secreting cells per conditions), IL-8 SI was calculated. A significant increase in SI indicates cellular up-regulation of that cytokine. The normalized data shows a significant cellular up-regulation of IL-8 only when MUTZ-LCs are present. This suggests that MUTZ-LCs are a requisite component of the IL-8 pro-inflammatory response. It does not necessarily indicate that MUTZ are producing all the IL-8 in the co-cultures, or even that they are the primary producers of IL-8 in co-culture. Based on the literature surrounding IL-8 secretion in moDCs and DC-like lines as predictive of sensitization, we suspect that MUTZ are the primary producers. However, it is well established that skin cells contribute to the pro-inflammatory response and cascade in sensitization both by cross-talk with immune cells and cytokine secretion[4-5]. We could confirm which cells are the primary producers of IL-8 by measuring RNA expression of inflammatory cytokines, and this is worth pursuit in future studies. These studies reinforced our belief that while MUTZ-LCs are likely the primary producers of IL-8, the presence of skin cells enhances detection and simulates a more physiologically relevant environment. We also determined that if

we are to pursue RNA expression of inflammatory cytokines, it would be reasonable to analyze expression in the MUTZ-LCs from the co-culture system.

As previously discussed, unreliable viability data calls the reliability of the SI into question. However, this limitation hardly affects our conclusions in these IL-8 contribution studies, as raw secretion also demonstrated significant IL-8 contribution only in the presence of MUTZ-LCs. The information gleaned from these studies should be considered in future pursuit of additional predictive metrics for this co-culture screening assay.

3.5 REFERENCES

1. Grone, A. (2002). "Keratinocytes and cytokines." Vet Immunol Immunopathol **88**(1-2): 1-12.
2. Luster, M. I., et al. (1995). "Role of keratinocyte-derived cytokines in chemical toxicity." Toxicol Lett **82-83**: 471-476.
3. Toebak, M. J., et al. (2006). "CXCL8 secretion by dendritic cells predicts contact allergens from irritants." Toxicol In Vitro **20**(1): 117-124.
4. Enk, A. H. (1997). "Allergic contact dermatitis: understanding the immune response and potential for targeted therapy using cytokines." Mol Med Today **3**(10): 423-428.
5. Hennen, J., et al. (2011). "Cross talk between keratinocytes and dendritic cells: impact on the prediction of sensitization." Toxicol Sci **123**(2): 501-510.

CHAPTER 4: THESIS CONCLUSIONS AND FUTURE WORK

4.1 SUMMARY OF FINDINGS

The high incidence of ACD compounded with the prevalence of sensitizing agents necessitates accurate screening methods to identify potential hazards. The major limitations of the currently established assays are decreased accuracy in prediction of pre- and pro-haptens and limited prediction of sensitizer potency. We previously developed an *in vitro* co-culture system of MUTZ-3-derived Langerhans cells, dermal fibroblasts, and HaCaT keratinocytes to address the limitation of pre-/pro-hapten prediction and mimic the components and prediction accuracy of a full-thickness skin model with LCs. The skin cells provide the metabolic components to activate pre- and pro-haptens and enhance the immune response.

In chapter 2 we established initial high accuracy of identification of skin sensitizers by the co-culture system (86%) as compared to MUTZ-LCs alone (45%) using IL-8 as a starting biomarker. We tested both systems with an expanded panel of chemicals, including sensitizers of every class and potency, non-sensitizers, and non-sensitizing irritants. From the initial IL-8 studies, it was immediately apparent that the co-culture system was superior to the MUTZ mono-culture in prediction of pre- and pro-hapten sensitizers. To increase this accuracy we analyzed the two culture systems' secretome for 27 inflammatory cytokines and used a support vector machine to identify and rank the most predictive metrics in combination. Using the top 3 biomarkers in combination, IL-8, MIP-1β, and GM-CSF, the co-culture system achieves an overall prediction accuracy of 91%, and a pre-/pro-hapten prediction accuracy of 90% with IL-8, RANTES, and GM-CSF. We demonstrated that a multi-metric signature offers superior prediction to single-metric. Classification trees were used to identify statistical thresholds to distinguish sensitizer from non-sensitizer and further classify potency. Using this approach, the co-culture system predicted no false negatives and correctly classified all weak/moderate sensitizers, achieving 83% accuracy overall in potency prediction. These findings demonstrate that the co-culture system

shows promise in mimicking the *in vivo* environment of sensitization and could thus be used to predict skin sensitizers. Furthermore, this system could potentially be used to model ACD *in vitro* for mechanistic studies. Identification of additional predictive metrics could establish further utility of this co-culture system as an *in vitro* approach to screen skin sensitizers across varying steps and sub-steps of the adverse outcome pathway of skin sensitization.

In chapter 3 the first steps to identify additional predictive metrics were taken. The cells from the co-culture system were treated in mono-cultures and co-cultures in every combination to assess under which conditions IL-8 was up-regulated in response to sensitizers. It was found that IL-8 was only up-regulated in the presence of MUTZ-LCs, indicating that the presence of MUTZ-LCs is requisite for a proper pro-inflammatory response in skin sensitization. This supports the motivation for our future work in identifying additional predictive metrics of skin sensitization.

4.2 FUTURE WORK

We are currently working to identify more predictive biomarkers to screen skin sensitizers and metrics for analysis of the ACD mechanism using our *in vitro* assay. The goal is to have one system that can be used not just for multi-metric secretome analysis but for multi-metric analysis across different steps of the adverse outcome pathway. Thus our co-culture system can be used as both a screening tool and a learning tool.

4.2.1 PCR ARRAY

Motivation

Both the GARD assay and the IL-8 Luc assay employ gene expression signatures as their biomarkers of sensitization. The GARD assay utilizes a 200 gene signature[1]; IL-8 Luc uses one[2]. As established in chapter 1, these assays have achieved moderate success in prediction of skin sensitizers; their main limitation is in pre- and pro-hapten prediction due to their lack of an incorporated metabolic component. Identifying a gene expression signature in our co-culture

assay could potentially offer even higher prediction accuracy in a manner that reduces the use of animal-derived monoclonal antibodies. While recombinant antibody technology that does not require the use of animals is on the rise, it is not yet mainstream. High cost has been a barrier to the adoption of recombinant antibodies (and even more so in multi-metric arrays), and few recombinant antibodies have been developed and are available for production[3]. Furthermore, as RNA expression is upstream of secretion, it could likely be measured at an earlier time-point, which might reduce potential problems with prolonged exposure to cytotoxic substances.

Materials and Methods

The co-culture system was set up as described in chapter 2. Four conditions were set up for the initial run: vehicle control, SDS at 250 μM, EU at 1000 μM, and IE at 1000 μM. MUTZ-LC pellets were generated from these conditions after 48 hours of treatment and flash-frozen in liquid nitrogen. The pellets were stored at -80°C until they were shipped to Qiagen for PCR and gene expression analysis. Based on our findings in chapters 2 and 3, the 84-gene human inflammatory cytokines and receptors array was chosen for analysis.

Progress

When we receive the RNA expression data generated by the PCR array we will use the SVM to identify the most predictive metrics. If the SVM succeeds in building a high-accuracy classification model with RNA expression, those markers could be used in addition to or in place of secretion markers. Additional Qiagen gene arrays, such as the dendritic and antigen presenting cell panel, the drug metabolism panel, or the oxidative stress panel, could be tested to identify additional predictive metrics from the other cells in the co-culture system. These panels could potentially shed light on different cellular roles in the sensitization process, as well, emphasizing the advantages of a co-culture system as a 2D model of skin sensitization.

4.2.2 MITOCHONDRIAL MARKERS OF SKIN SENSITIZATION

Motivation

The skin microenvironment produces danger signals, including reactive oxygen species (ROS), uric acid, nitric oxide, and hyaluronic acid fragments, in the event of cellular stress. The role of ROS in eliciting an allergic response is not clear. Most ROS are generated during mitochondrial electron transport, but studies have shown that ROS produced in response to allergens is sometimes cytosolic, and different allergens may preferentially trigger ROS production from different intracellular sources[4-6]. This brings the mitochondrial role in sensitization into question. Mitochondria are key regulators of cellular homeostasis and metabolism, and metabolic stress can trigger mitochondrial fragmentation and mitophagy[7]. Understanding the mitochondrial response to sensitizers and non-sensitizing irritants may shed light on mechanisms of chemical metabolism and offer other metrics to explore for screening assays.

Materials and Methods

Our preliminary studies involve mitochondrial imaging of the dermal fibroblasts from our co-culture method. FBs were chosen for their convenience; they are large and adherent and therefore a model cell type for development of fluorescent imaging methods. Dermal fibroblasts were cultured as described in chapter 2. For high-resolution microscopy, oil immersion methods were utilized. To improve adherence to the non-cell culture-treated #1 borosilicate chambered coverglass, the glass was coated with Poly-D-Lysine (Sigma) for 3 hours prior to seeding. Cells were seeded at a density of $2.4x10^4$ cells/mL and adhered overnight. The following day the fibroblasts were stained with 50 nM MitoTracker green (Thermo Fisher Scientific) for 15 minutes and washed 3x with 1x PBS. The cells were then treated with a panel of sensitizers (EU, IE, HQ, DNCB) and non-sensitizers (XYL, LA, SDS) and incubated for 1-4 hours at 37°C and 5% CO_2. Images were taken with an Olympus IX81 confocal microscope using differential interference contrast (DIC) and GFP (green) channels.

Preliminary Results

We developed an assay to analyze mitochondrial changes in response to sensitizers using dermal fibroblasts as the model cell. This assay allows for real-time analysis, that is, continuous analysis over time. We tested 3 non-sensitizers and 4 sensitizers and found that both LA and XYL appeared similar to vehicle, while SDS induced marked mitochondrial rounding. Weak and moderate sensitizers EU and IE demonstrated mitochondrial clustering towards the nucleus, and the strong sensitizers HQ and DNCB killed the fibroblasts within hours (**Figure 4.1**). Cell death in response to strong sensitizers was apparent by sight; very few cells remained attached to the plate, and those that remained were shriveled and small in comparison to live, healthy cells (image not shown). Further study is necessary to confirm mitochondrial morphological trends in response to sensitizers and non-sensitizers.

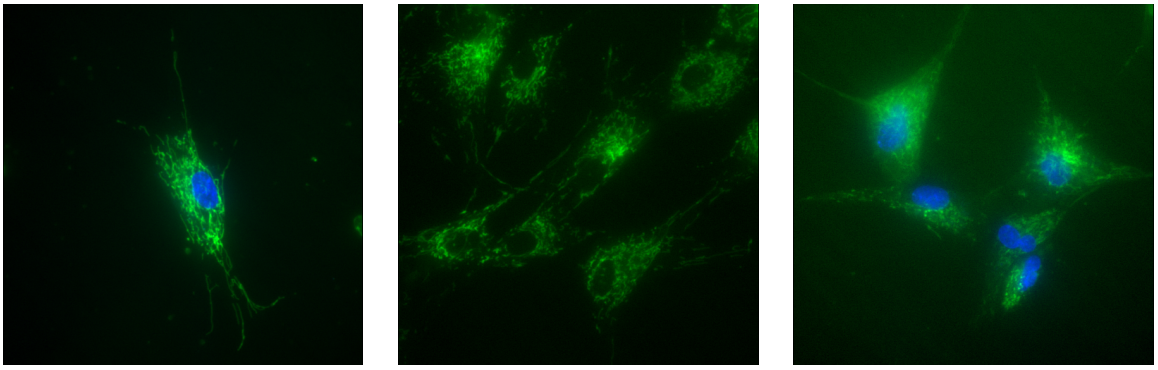

Figure 4.1 Representative images of changes in mitochondrial morphology. Vehicle image (left) demonstrates healthy, elongated, aligned mitochondria. SDS image (middle) shows rounded mitochondria, indicating fission and/or cellular stress. IE image (right) shows mitochondrial clusters aggregated around nuclei.

These preliminary findings raise questions regarding the 48-hour time point chosen for our co-culture system, as there are detectable mitochondrial changes and apparent cellular stress within 1-4 hours of treatment. It is possible that the chemicals are metabolized at different rates and thus activate cellular stress at different rates. The reliability of Alamar Blue comes into question, as well, as the concentrations of HQ and DNCB used in these studies were very evidently cytotoxic. However, it is also possible that keratinocytes rapidly detoxify these chemicals in the co-culture

system, thus reducing cytotoxicity. This will require further study in mono-cultures and co-cultures of the cell types included in our co-culture system. Additional image analysis of co-culture experiments demonstrated significant cell death in conditions treated with high concentrations of a few of our selected moderate sensitizers and strong/extreme sensitizers after 48 hours that does not align with our findings from Alamar Blue (data not shown). Therefore it might be advisable to optimize our Alamar Blue analysis by comparing the results from various time points to an exclusion dye, such as propidium iodide or trypan blue.

Looking Ahead

This mitochondrial imaging assay will be extended to include keratinocytes and MUTZ-LCs in mono-culture and eventually the co-culture system to assess the potential of mitochondrial morphology markers as predictive metrics of chemical metabolism or sensitization. Other imaging techniques with a mitochondrial activity stain may offer additional metrics. However, these techniques are not without limitations. Perturbation of mitochondrial function during apoptosis can lead to oxidative stress, which has been shown to induce increased fluorescence in certain mitochondrial stains[8]. Optical scatter imaging has offered opportunity for label-free imaging of mitochondria; if mitochondrial markers prove to be predictive of sensitization, this label-free method could be implemented with our co-culture system[9]. Automating the imaging process or image analysis could present a high-throughput and efficient screening method. Studies comparing cytosolic and mitochondrial ROS levels in response to sensitizers and irritants have shown some promise; this too, merits further study[6].

Ultimately, these studies may bring additional predictive biomarkers of sensitization to light while enhancing our current understanding of the mitochondrial roles in chemical metabolism and allergic contact dermatitis. Combining the findings of these future studies with the results of this thesis work may prove our co-culture system to be a valuable, multi-purpose tool in the advancement of the field of skin allergy.

4.3 REFERENCES

1. 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.
2. Takahashi, T., et al. (2011). "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.
3. Groff, K., et al. (2015). "Modern affinity reagents: Recombinant antibodies and aptamers." Biotechnol Adv **33**(8): 1787-1798.
4. Mehrotra, P., et al. (2005). "Differential regulation of free radicals (reactive oxygen and nitrogen species) by contact allergens and irritants in human keratinocyte cell line." Toxicol Mech Methods **15**(5): 343-350.
5. Galbiati, V., et al. (2014). "Role of ROS and HMGB1 in contact allergen-induced IL-18 production in human keratinocytes." J Invest Dermatol **134**(11): 2719-2727.
6. Corsini, E., et al. (2013). "Role of oxidative stress in chemical allergens induced skin cells activation." Food Chem Toxicol **61**: 74-81.
7. Toyama, E. Q., et al. (2016). "Metabolism. AMP-activated protein kinase mediates mitochondrial fission in response to energy stress." Science **351**(6270): 275-281.
8. Pendergrass, W., et al. (2004). "Efficacy of MitoTracker Green and CMXrosamine to measure changes in mitochondrial membrane potentials in living cells and tissues." Cytometry A **61**(2): 162-169.
9. Boustany, N. N., et al. (2002). "Calcium-induced alterations in mitochondrial morphology quantified in situ with optical scatter imaging." Biophys J **83**(3): 1691-1700.