Predicting Skin Sensitization Using 21st Century Toxicology

N Kleinstreuer[1], J Strickland[1], D Allen[1], W Casey[2]

[1]ILS/NICEATM, RTP, NC, USA; [2]NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA

Introduction

- Regulatory, economic, and animal welfare concerns are driving international efforts to eliminate animal use for skin sensitization testing.
- The U.S. Tox21 and ToxCast high-throughput screening (HTS) projects include assays with endpoints that are relevant to skin sensitization and map to the adverse outcome pathway (AOP) for skin sensitization. These assays have been used to test hundreds of potential skin sensitizers.
- We identified 60 chemicals from the NICEATM LLNA database (36 LLNA positives and 24 LLNA negatives) with ToxCast assay data and used the HTS data to build a model to predict LLNA results.

Eliminating Animal Use for Skin Sensitization Testing

- National and international regulatory authorities require testing of pesticides, personal care products, and other chemical products to assess their potential to cause allergic contact dermatitis (ACD).
- The murine local lymph node assay (LLNA) is the current preferred animal test to identify skin sensitizers, or substances with the potential to cause ACD. However, regulatory, economic, and animal welfare concerns are driving international efforts to eliminate animal use for this purpose.
- In the U.S., the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) evaluates alternative test methods for assessing the potential of chemicals to cause ACD.
    - NICEATM is collecting and curating high-quality mouse and human skin sensitization data (**Table 1**).
    - These data will support the development of mechanistic models and *in vitro* and *in silico* testing approaches to identify potential skin sensitizers.

Table 1 NICEATM *In Vivo* Skin Sensitization Databases Currently Under Development

Database	Number of Chemicals	Number of Studies per Chemical [range]	% Positive (any study)	Tox21 Chemical Overlap	ToxCast Chemical Overlap
LLNA	668	[1-43]	65%	273	60
Human	127	[1-8]	59%	79	24

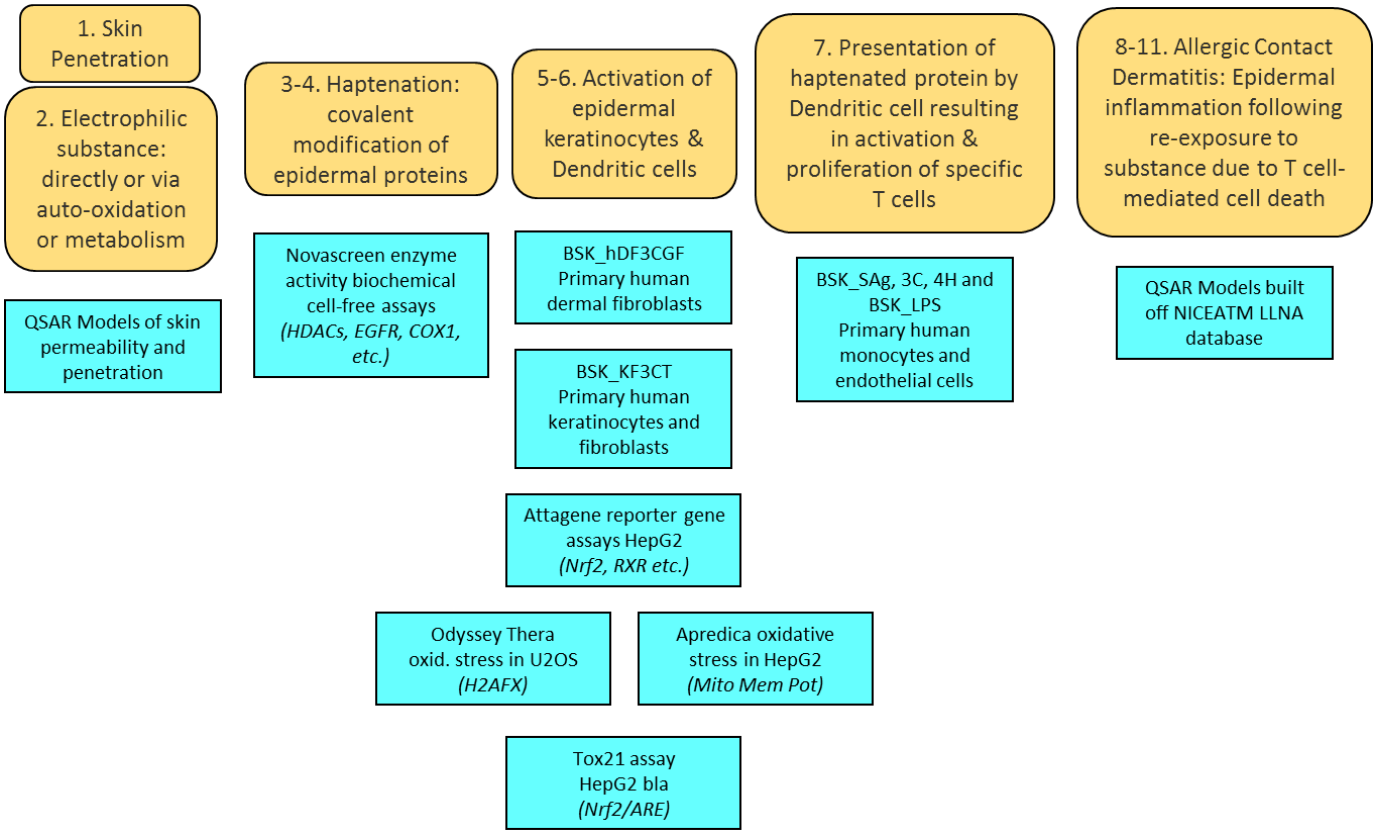
U.S. Programs for High-Throughput Chemical Testing

- The U.S. Tox21 and ToxCast high-throughput screening projects include assays with endpoints that are relevant to skin sensitization. These assays have been used to test hundreds of potential skin sensitizers.
  - The 10,000 chemicals in the Tox21 library have been tested using over 60 Tox21 assays, including assays to measure induction of IL-8, TNFα, and Nrf-2. These data are still undergoing analyses (Tice et al. 2013).
  - The 1047 chemicals in the ToxCast Phase I and II libraries have been screened using over 700 ToxCast assays, including endpoints in primary human skin cells.

Alignment of ToxCast Assays with the Skin Sensitization Adverse Outcome Pathway

- The Organisation for Economic Co-operation and Development (OECD) has developed an adverse outcome pathway (AOP) for skin sensitization initiated by covalent binding to proteins (OECD 2012). The AOP identifies key events that occur after exposure to certain types of skin sensitizers that result in the development of ACD (**Figure 1**, yellow boxes).
- The ToxCast assay portfolio includes *in vitro* assays that use human cell types involved in development of ACD, such as keratinocytes and monocytes. These and other ToxCast assays measure many biological processes relevant to ACD. The green boxes in **Figure 1** show how ToxCast assays align to the steps of the skin sensitization AOP.

Figure 1 Preliminary Mapping of *In Vitro* Assays and *In Silico* Models to the Skin Sensitization AOP



Abbreviations: BSK = BioSeek; LLNA = murine local lymph node assays; oxid. = oxidative; QSAR = quantitative structure–activity relationship.

Developing a Model to Use ToxCast Data to Predict LLNA Results

- ToxCast assay data from 60 chemicals (36 LLNA positives and 24 LLNA negatives) were used to build a model to predict LLNA results for chemicals in the NICEATM LLNA database (**Table 1**).
- A random forest (RF) model was created using ToxCast assay results as descriptors and LLNA results as an endpoint.
    - RF is an ensemble machine learning technique based on randomized decision trees (Breiman 2001). The outputs of all trees are aggregated to obtain one final prediction.
    - Each tree is grown as follows:

        (i) A bootstrap sample is performed on the entire set of N compounds to form a training set for the current tree. The compounds omitted from the training set are placed in the out-of-bag (OOB) set (size ~ N/3).

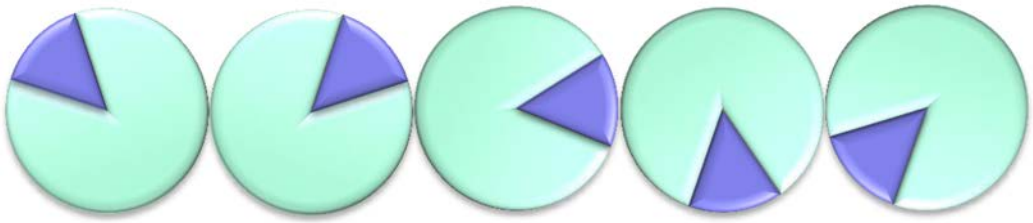
(ii) The best split among the randomly selected descriptors from the entire pool at each node is chosen.

(iii) Each tree is grown to the largest possible extent without pruning. The predicted classification values are defined by majority voting, and each tree predicts values for only those compounds in the OOB set.

The final model is chosen by the lowest error for prediction of the OOB set.

- Five-fold cross-validation was performed by dividing the data independently into training sets (80% of data) and test sets (20% of data) five times, as illustrated in **Figure 2.**

Figure 2 Five-Fold Cross-Validation of RF Model



Results

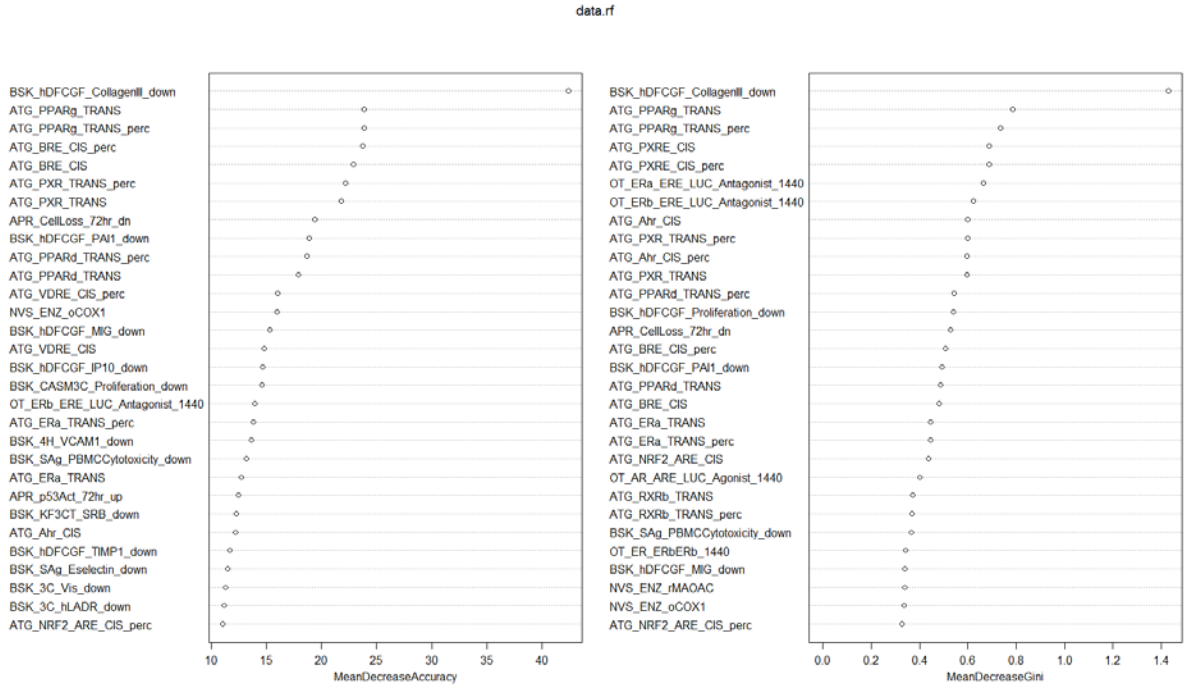
- Each iteration produced an RF model with features (ToxCast assays) ranked by variable importance in predicting the LLNA results.
- The cross-validated model produced an average specificity of 73% and an average sensitivity of 88%, with a negative predictive value of 69% and a positive predictive value of 92%.
- To build the final model and to avoid overfitting due to dimensionality concerns, the top 100 features across all model iterations based on their variable importance scores were used. The final model was retrained on the entire set of 60 chemicals with ToxCast and LLNA data.
- The final OOB error estimate was 20%, representing the accuracy of the model (80% [48/60]) against all chemicals when they appear in the external test sets. **Table 2** shows the confusion matrix and the correct classification rates for sensitizers and nonsensitizers.

Table 2 Confusion Matrix for Final RF Model: External Test Set Performance

	Nonsensitizer	Sensitizer	Classification Error
Nonsensitizer	20	4	0.17
Sensitizer	8	28	0.22

- **Figure 3** shows the 30 most important ToxCast assays for predicting LLNA results from the final RF model. Assays are ranked by the mean decrease in accuracy resulting when that feature is removed (left plot), and the mean decrease in Gini score, a measure of node purity and feature relevance (right plot).

Figure 3 RF Model: Variable Importance Plot

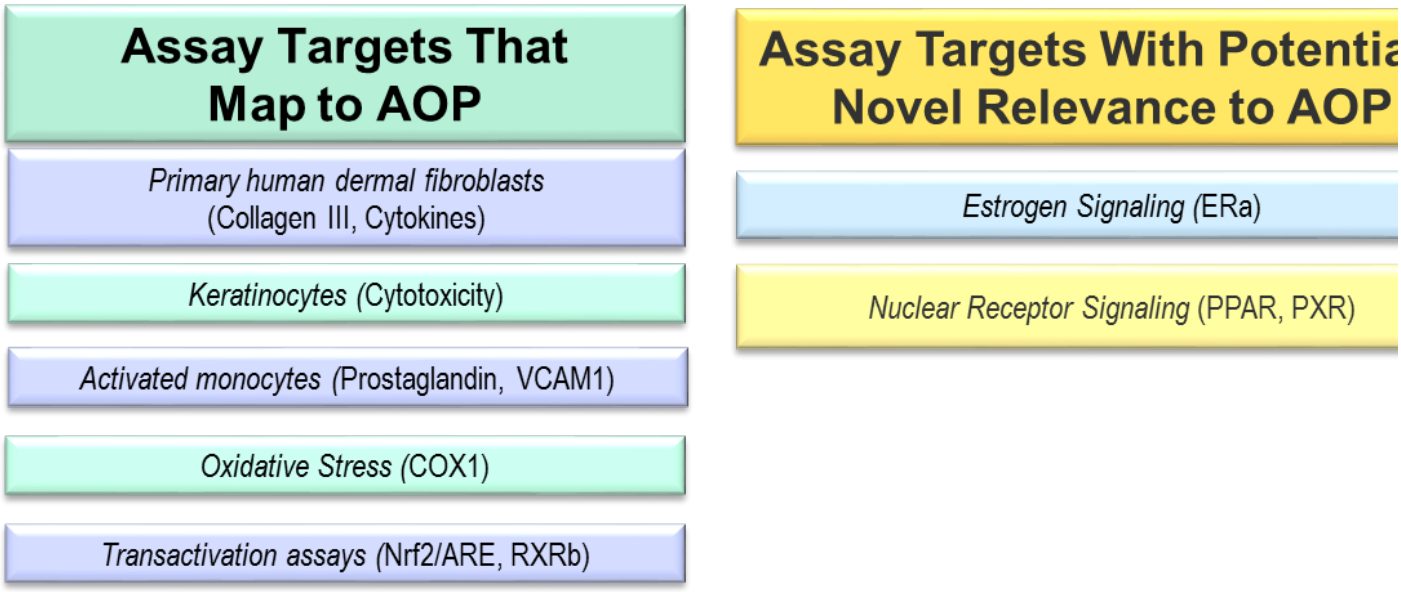


- The ToxCast assay with the highest variable importance in the model was Collagen Type 3 (Col III) regulation in primary human dermal fibroblasts (BSK_hDFCGF_CollagenIII_down, the first assay listed in each plot).
  - Out of the 36 positive chemicals, 23 produced at least 2-fold downregulation of Col III expression in this assay.
  - Only 3 of the negative chemicals showed this activity.

- When retrained on the entire data set and applied to predict categorizations of the training set, the RF model using data from the assays shown in **Figure 3** had an accuracy of 95% (57/60).
- The model failed to correctly classify three chemicals: diethyl sulfate, methyl methanesulfonate, and dimethyl sulfate.
    - These compounds, all classified by LLNA data as moderate or strong sensitizers, were largely inactive across the ToxCast assays. The model therefore incorrectly predicted them to be nonsensitizers.
    - The chemicals are structurally similar; hybrid models that incorporate molecular descriptors may assist in identifying them as sensitizers.

Discussion

Figure 4 Summary of Most Important ToxCast Assays for Predicting LLNA Results



Abbreviations: AOP = adverse outcome pathway.

Most Important RF Model Features Map to AOP

- Despite the limited overlap (n = 60 chemicals) between the ToxCast chemical library and the NICEATM LLNA database, the unsupervised modeling process identified a number of ToxCast assays (**Figure 3**) that map to the skin sensitization AOP based on biological relevance (**Figure 4,** left panel, and **Figure 1**).
  - **Assays using primary human dermal fibroblasts:**
    - Fibroblast proliferation is controlled by lymphocyte signaling and is affected by specific antigenic challenges.
    - Fibroblast proliferation is involved in oxidative stress signaling in skin sensitization (Wondrak et al. 2003).
    - Collagen in chronically inflamed dermal tissue has altered biochemical characteristics and functions (Hirota et al. 2003).
  - **Assays using keratinocytes:**
    - Keratinocytes contribute to skin sensitization by producing cytokines, for example interleukins and TNFalpha, in response to cellular injury or cytotoxicity.

  - Various chemokines are elevated in atopic dermatitis via enhanced production by keratinocytes and self-perpetuating inflammatory mechanisms.

- **Assays using activated monocytes:**
  - Monocyte signaling regulates activation and proliferation of T cells.
  - Endothelial adhesion molecules such as VCAM1 can be upregulated by CD40L on activated T cells and are critical for memory T-cell infiltration.
  - Prostaglandin signaling promotes allergic skin inflammation in response to cutaneous exposure to antigen via T-cell receptor mediated prostaglandin responsive chemotaxis (He at al. 2010).
- **Oxidative stress/transactivation assays:**
  - Oxidative stress activates transcription factors and signaling pathways, including NF-kB and p38 MAPK, leading to the release of cytokines and chemokines.
  - Reactive oxygen species serve as messengers mediating cellular responses and resulting in immune cell activation (Corsini et al. 2013).

The Model Identified Some Potential Novel Targets

- A number of novel targets were identified that do not map to the current AOP (**Figure 4,** right panel).
  - **Estrogen signaling assays**:
    - Estrogens affect skin physiology and pathophysiology; human skin fibroblasts express estrogen receptor alpha and beta (Haczynski et al. 2002).
  - **Nuclear receptor signaling assays**:
    - Human and rodent skin cells express all three PPAR isotypes, which play an important role in inflammatory responses and signaling (Sertznig and Reichrath 2011).
    - PXR, expressed in skin and especially in proliferating keratinocytes, is known to regulate oxidative stress and to control cell proliferation.

Conclusions

- The AOP for skin sensitization provides a theoretical basis to identify and develop *in silico* models and *in vitro* assays that cover the relevant biology associated with this common human hazard.
- Well-curated toxicological reference databases combined with large HTS datasets can be used to develop models for testing strategies that minimize the use of animals in regulatory testing.
- Using such a model, we identified the most important ToxCast assays for predicting LLNA results. The model predicts LLNA results with 80% accuracy, and includes assays representing known AOP targets as well as targets outside of the current AOP.
- Future goals include:
    - Collection and curation of additional *in vivo* and *in vitro* data
    - Developing multiclass and continuous models to predict sensitization potential and potency
    - Using human data to develop models to support the design of a testing strategy that will accurately predict clinical responses
    - Collaborating with QSAR researchers to develop hybrid models that incorporate structural descriptors and *in vitro* data for improved predictivity

References

Breiman L. 2001. Machine Learning 45: 5–32.

Corsini E, Galbiati V, Nikitovic D. Tsatsakis AM. Food Chem Toxicol 61: 74–81.

Haczynski J, Tarkowski R, Jarzabek K, et al. 2002. Int J Mol Med 10: 149–153.

He R, Oyoshi MK, Wang JY, Hodge MR, Jin H, Geha RS. 2010. J Allergy Clin Immunol 126: 784–790.

Hirota A, Ebihara T, Kusubata M, et al. 2003. J Invest Dermatol 121: 1317–1325.

OECD. 2012. OECD Series on Testing and Assessment No. 168. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Assessment. Paris:OECD Publishing. Available: http://www.oecd.org/env/ehs/testing/adverse-outcome-pathways-molecular-screening-and-toxicogenomics.htm [accessed 2 Dec 2013]

Sertznig P, Reichrath J. 2011. Dermatoendocrinol 3: 130–135.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Environ Health Perspect 121: 756–765.

Wondrak GT, Roberts MJ, Cervantes-Laurean D, Jacobson MK, Jacobson EL. 2003. J Invest Dermatol 121: 578–586.

Acknowledgements

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by ILS under NIEHS contracts N01-ES 35504 and HHSN27320140003C.

The views expressed above do not necessarily represent the official positions of any Federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.

A summary of NICEATM activities at the Ninth World Congress is available on the National Toxicology Program website at http://ntp.niehs.nih.gov/go/41583.