Regulatory Toxicology and Pharmacology 94 (2018) 276–282

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Mechanism-informed read-across assessment of skin sensitizers based on SkinSensDB

Chun-Wei Tung[a,b,c,d,*], Chia-Chi Wang[a,b,c,d], Shan-Shan Wang[a]

[a] School of Pharmacy, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
[b] PhD Program in Toxicology, Kaohsiung Medical University, Kaohsiung 80708, Taiwan
[c] Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung 80708, Taiwan
[d] National Institute of Environmental Health Sciences, National Health Research Institutes, Miaoli County 35053, Taiwan

ARTICLE INFO

Keywords:
Adverse outcome pathway
Skin sensitizer
Read-across
SkinSensDB

ABSTRACT

Integrative testing strategies using adverse outcome pathway (AOP)-based alternative assays for assessing skin sensitizers show the potential for replacing animal testing. However, the application of alternative assays for a large number of chemicals is still time-consuming and expensive. In order to facilitate the assessment of skin sensitizers based on integrative testing strategies, a mechanism-informed read-across assessment method was proposed and evaluated using data from SkinSensDB. First, the prediction performance of two integrated testing strategy models was evaluated giving the highest area under the receiver operating characteristic curve (AUC) values of 0.928 and 0.837 for predicting human and LLNA data, respectively. The proposed read-across prediction method achieves AUC values of 0.957 and 0.802 for predicting human and LLNA data, respectively, with interpretable activation statuses of AOP events. As data grows, a better prediction performance is expected. A user-friendly tool has been constructed and integrated into SkinSensDB that is publicly accessible at http://cwtung.kmu.edu.tw/skinsensdb.

1. Introduction

Chemicals acting as haptens could form a complex with protein molecules that may trigger T cell-mediated immune reactions and lead to allergic contact dermatitis (Karlberg et al., 2008), the second most common occupational illness accounting for 10–15% of all occupational disease worldwide (Basketter et al., 2015). The assessment of skin sensitization is traditionally based on *in vivo* assays such as murine local lymph node assay (LLNA) and guinea pig maximization test (GPMT). Due to the ban on animal testing for cosmetic ingredients in 2013 by the European Union and 3Rs (Replacement, Reduction and Refinement) principle, the paradigm for assessing skin sensitizers has been gradually shifted to alternative non-animal assays (Mehling et al., 2012; Vandebriel and van Loveren, 2010). Previous studies on the mechanism of skin sensitization have formulated the process as an adverse outcome pathway (AOP) including four key events of protein binding, keratinocyte activation, dendritic cell activation and T-cell activation (OECD, 2012).

Several alternative assays have been widely adopted for assessing the potential of skin sensitization including Direct Peptide Reactivity Assay (DPRA) and Peroxidase Peptide Reactivity Assay (PPRA) (Gerberick et al., 2009, 2004) for protein binding, KeratinoSens (Emter et al., 2010; Natsch and Emter, 2008) and LuSens (Bauch et al., 2012) for keratinocyte activation, and h-CLAT (Ashikaga et al., 2002; Sakaguchi et al., 2006) for the activation of dendritic cells, respectively. While each assay provides a good accuracy, a few integrated testing strategies based on multiple alternatives methods have been proposed to further improve the prediction performance on murine and human data. Integrated testing strategies were firstly developed as a majority vote or tiered approach integrating several assays for achieving highest prediction performance (Bauch et al., 2012; Natsch et al., 2013; van der Veen et al., 2014). Jaworska et al. proposed a Bayesian integrated testing strategy capable of predicting 21 test compounds with a high accuracy by integrating *in chemico*, *in vitro* and *in silico* methods (Jaworska et al., 2013). A Bayesian integrated testing strategy based on only validated assays was subsequently reported (Jaworska et al., 2015). In addition to probabilistic models, machine learning algorithms of artificial neural network, logistic regression and support vector machine have also been developed for integrating multiple assays (Hirota et al., 2015,2013; Strickland et al., 2017; Tsujita-Inoue et al., 2014). In contrast to complex machine learning algorithms, Urbisch et al. developed a very simple '2-out-of-3' approach based on a majority vote

* Corresponding author. Room N515, First-Teaching Building, No. 100, Shih-Chuan 1st Rd., Kaohsiung 807, Taiwan.
E-mail address: cwtung@kmu.edu.tw (C.-W. Tung).

https://doi.org/10.1016/j.yrtph.2018.02.014
Received 11 December 2017; Received in revised form 14 February 2018; Accepted 22 February 2018
Available online 24 February 2018
0273-2300/

from assays representing three key events with accuracies of 90% and 79% for predicting human and LLNA data (Urbisch et al., 2015). The potency of skin sensitizers can also be evaluated by integrated testing strategies (Nukada et al., 2013; Tsujita-Inoue et al., 2014). While the '2-out-of-3' approach provides a balanced prediction, the combination of KeratinoSens and h-CLAT may be utilized for high sensitivity assessment (Otsubo et al., 2017). Roberts et al. further developed a decision tree-based method utilizing only DPRA and h-CLAT assays whose accuracy was higher than that of '2-out-of-3' approach based on 271 chemicals (Roberts and Patlewicz, 2017). Although integrated testing strategies are useful for identifying skin sensitizers, the experimental assessment of AOP assays for a huge number of chemicals is still time-consuming and expensive. It is therefore desirable to develop computational methods leveraging existing experimental data for prioritizing potential skin sensitizers.

This study presents a novel method combining integrated testing strategies for mechanism-based assessment of skin sensitizers and a similarity-based read-across method for assessing the activation of AOP events using data from SkinSensDB (Wang et al., 2017). SkinSensDB is a manually curated database for AOP assays consisting of > 700 unique chemicals and > 5000 associated AOP assay values providing a useful resource for read-across assessment of skin sensitizers. Two models of the majority vote (2-out-of-3) (Otsubo et al., 2017) and decision tree (Roberts and Patlewicz, 2017) have been implemented and integrated into SkinSensDB for predicting LLNA and human responses. The optimal similarity threshold for read-across assessment and corresponding prediction performance have been thoroughly evaluated. The proposed method is expected to be useful for mechanism-informed read-across assessment of skin sensitizers.

2. Results

2.1. Performance of integrative testing strategy models

Two integrative testing strategy models of the majority vote (2-out-of-3) (Otsubo et al., 2017) and decision tree (Roberts and Patlewicz, 2017) have been implemented and integrated into SkinSensDB (Wang et al., 2017). In contrast to previous studies whose performance evaluation was based on chemicals collected from a few datasets, SkinSensDB consisting of assay values curated from many literatures can be utilized to evaluate the robustness of the models. In this study, the prediction performance for both LLNA and human data was evaluated.

Firstly, chemicals without missing data for all three events of protein binding, keratinocyte activation, and activation of dendritic cells were extracted from SkinSensDB. After removing chemicals without LLNA or human data, 111 (80 sensitizers and 31 non-sensitizers) and 66 (46 sensitizers and 20 non-sensitizers) chemicals with LLNA and human data, respectively, were utilized for evaluating the prediction performance of the two models. For a given chemical, the final result representing the activation status for each AOP event was determined by a majority vote based on the curated experimental results. The two models were then applied to predict LLNA and human data based on the activation statuses of AOP events.

Table 1 shows the detailed prediction performance of the two integrative testing strategy models. Based on the area under the receiver operating characteristic curve (AUC), the majority vote model with AUC values of 0.928 and 0.837 is slightly better than the decision tree model with AUC values of 0.922 and 0.805 for predicting human and LLNA data, respectively. Generally, the decision tree model achieves a higher sensitivity with a lower specificity due to its relatively strict criteria of non-sensitizers that both events should be negative. According to our results, the integrative testing strategy models are more predictive of human data and both provide reasonably good performance. The activation status for each event and predicted skin sensitization potential are available in Table S1 and S2.

Table 1
Prediction performance of integrative testing strategy models using chemicals without missing data for all three events.

Endpoint	Models	Sensitivity	Specificity	Accuracy	AUC
Human data	Majority vote	89.13%	90.00%	89.39%	0.928
Human data	Decision tree	95.65%	70.00%	87.88%	0.922
LLNA data	Majority vote	86.25%	67.74%	81.08%	0.837
LLNA data	Decision tree	92.50%	48.39%	80.18%	0.805

2.2. Mechanism-informed read-across assessment of skin sensitizers

A novel AOP-based read-across assessment method for skin sensitizers has been developed. For a given chemical, the activation status for three AOP events will be firstly determined and integrative testing models will be subsequently applied to predict the final outcome. To evaluate the prediction performance of the proposed method, chemicals without complete data for corresponding alternative assays were firstly extracted from SkinSensDB. For the majority vote model, chemicals with any missing data for three events of protein binding, keratinocyte activation and activation of dendritic cells were included in the read-across evaluation. In contrast, chemicals with any missing data for protein binding and activation of dendritic cells were included for evaluating the decision tree model. For predicting LLNA data, a total of 372 and 348 chemicals were evaluated for the majority vote and decision tree models, respectively. For predicting human data, the dataset consists of 55 and 48 chemicals for evaluating the majority vote and decision tree models, respectively. Firstly, for each chemical, the missing assay data will be replaced by a predicted value using a 1-nearest neighbor algorithm, i.e., experimental results from the most similar chemical from SkinSensDB. Subsequently, the two integrative testing strategy models were applied to predict LLNA and human data.

To evaluate the effects of chemical similarity on prediction performance, the two models were applied to predict chemicals with a similarity score greater than or equal to a given threshold Tc∈{0, 0.1, …, 0.9}. Please note that performance evaluation using a threshold of 0.8 was not conducted for human data due to the small number of chemicals (≤15) filtered by the thresholds. Performance evaluation with only a few chemicals could be biased. The prediction performance is shown in Fig. 1. As expected, the two models are more predictive of the human data rather than LLNA data that are consistent with the results for chemicals without missing data (Table 1). Also, a higher similarity threshold gives better AUC performance with one exception (Tc = 0.8 for predicting LLNA data). According to our results, a threshold greater than or equal to 0.7 is suggested for read-across assessment of skin sensitizers. The best model for read-across prediction of human skin sensitization is the majority vote model with an AUC of 0.957 (Tc = 0.8), compared with the decision tree model with an AUC of 0.700 (Tc = 0.7). Notably, the decision tree model was evaluated using a smaller dataset due to the ignorance of missing data for keratinocyte activation that could result in a lower performance. As for predicting LLNA data, the highest AUC values of 0.802 and 0.771 for the majority vote and decision tree models were achieved using a similarity threshold Tc = 0.9. As the data curated in SkinSensDB grows, the proposed method is expected to be more accurate. Detailed information of the evaluated chemicals, their read-across prediction of outcomes and similarities for each event, and predicted skin sensitization potential is available in Table S3, S4, S5 and S6.

To compare with the traditional way of similarity-based prediction of skin sensitizers, the same datasets without missing assay information consisting of 111 and 66 chemicals with LLNA and human data were utilized as training datasets. The 1-nearest neighbor algorithm was applied to predict 372 and 55 chemicals in the validation sets. As shown in Fig. 1, the direct prediction of human data with the highest AUC value of 0.700 (Tc = 0.7) provides the same prediction performance as