News Release

October 15,2025
Pigeon Corporation

Paper Published in Online Edition of SCI-listed International Academic Journal *Peerj Computer Science*

Sensitization Properties in Cosmetic Compounds

PIGEON MANUFACTURING (SHANGHAI) CO., LTD. (Location: Qingpu District, Shanghai; Chairman: Ken Kaku), a subsidiary of Pigeon Corporation (Headquarters: Tokyo; President: Ryo Yano) responsible for advancing baby skin research within the Pigeon Group, is pleased to announce that its research paper regarding the "Detection and Evaluation of Skin Sensitizing Components using Machine Learning (AI)" has been published in the online edition of the SCI-indexed international academic journal, *PeerJ Computer Science*. The research was conducted at the company's Pigeon Skincare Research Center.

In its recent research, PIGEON MANUFACTURING (SHANGHAI) CO.,LTD is exploring new methods for identifying latent allergens included in skincare compounds. When certain chemical substances come into contact with the skin, the body interprets them as a foreign substance, and subsequent contact with the same substance can make allergic reactions more likely. This is called *skin sensitization*. The animal testing formerly conducted to test for skin sensitization properties is now restricted, increasing the need for accurate non-animal testing methods. PIGEON MANUFACTURING (SHANGHAI) CO.,LTD has a policy of not performing animal testing to confirm at the product development stage whether skin sensitization occurs. As an alternative method, it developed a system combining two kinds of skin cells: HaCaT keratinocytes and THP-1 cells. These cells were prepared to use machine learning technology. This led to the development of a better method for testing for skin sensitization properties, with no need to rely on animal testing.

Summary of research paper

Background

To improve the precision of identifying allergens in cosmetics, a system was developed combining two kinds of skin cells: HaCaT keratinocytes and THP-1 cells. These cells were prepared to be usable in machine learning technology. The objective was to establish a superior method for evaluating skin sensitization properties without relying on animal testing.

Methods

First, the aforementioned cell system combining two types of cells, HaCaT keratinocytes and THP-1 cells, was exposed to various chemical substances. By confirming cell health, we discovered safe exposure levels. Next, we extracted RNA from THP-1 cells and analyzed it using the genetic analysis technology RNA-Seq. From this analysis, we identified the genes that indicated different activation levels when exposed to chemical substances. After that, we processed and analyzed this data using various machine learning methods and constructed a model that can predict whether a substance will cause skin sensitization. Eight different

machine learning methods were tested and compared.

Results

This unified approach was proven effective. Several machine learning models showed high precision in predicting skin sensitization properties. The most superior models, “Random Forest” and “voom-based diagonal quadratic discriminant analysis (voomDQDA)” reached 100% precision in testing. Other models, such as “Support Vector Machine (SVM)” and “voom-based nearest shrunken centroids (voomNSC)” also showed extremely high performance at 96.7%. These results indicate that this method is able to successfully differentiate between substances that cause sensitization properties and those that do not.

Discussion

This research shows that it is possible to create a more accurate and comprehensive lab test for skin sensitization properties by combining a skin culturing system, advanced genetic analysis (RNA-Seq), and machine learning. This insight is a step forward in evaluating the safety of cosmetics and could prove useful in reducing the need for animal testing. The researchers discovered that the data preparation methods and genetic feature selection methods affect the model’s predictive precision. Issues such as sample size and model interpretation remain, but this research offers valuable insight for future development in this field.

Fig. 1. Using THP-1 cells to evaluate the cell toxicity of 15 types of test compound at concentrations between 1000 μg/mL and 3.9μg/mL. The cells’ survival rate was evaluated using a CCK-8 assay 48 hours after processing.

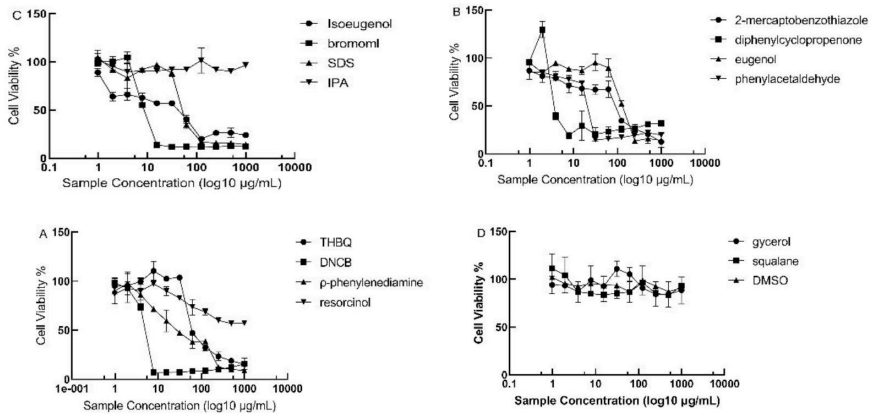
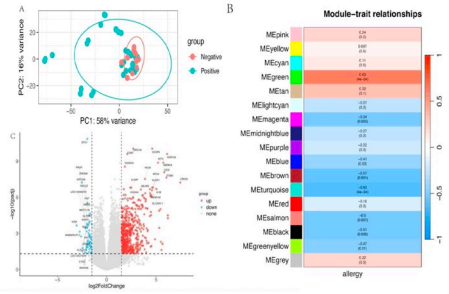
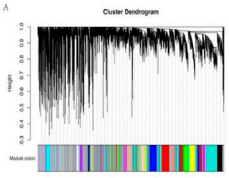


Fig. 2. Analysis of different expressed genes. (A) PCA of allergy and non-allergy group. (C) DEG volcano plot.

Fig. 3. WGCNA analysis. (A) Gene dendrogram obtained via clustering differences. (B) Relationship between consensus module feature genes and allergies.





ML classifier (lasso-logcpm)	1st Method with Vars(100) genes		2nd Method with DEG(100) genes		3rd Method with WGCNA(100) genes	
	Train accuracy	Test accuracy	Train accuracy	Test accuracy	Train accuracy	Test accuracy
Random Forest (RF)	0.833	0.905	0.867	0.952	0.862	0.905
K-Nearest neighbor(KNN)	0.833	0.667	0.867	0.810	0.867	0.857
Support Vector Machine (SVM)(vomRadial/Poly)	0.867/0.90	0.905/0.952	0.90/0.90	0.905/0.905	0.896/0.866	0.952/0.857
Nearest Shrunken Centroids (NSC)/vomNSC	0.867/0.8333	0.952/0.619	0.867/0.967	0.905/0.810	0.833/0.833	0.619/0.619

Table Comparison of Training and Test Accuracies for Three Feature Selection Methods Using different Processing

ML classifier (deseq-vst)	1st Method with Vars(100) genes		2nd Method with DEG(100) genes		3rd Method with WGCNA(100) genes	
	Train accuracy	Test accuracy	Train accuracy	Test accuracy	Train accuracy	Test accuracy
Random Forest (RF)	0.871	0.810	0.867	0.857	0.862	0.905
K-Nearest neighbor(KNN)	0.867	0.810	0.867	0.857	0.867	0.857
Support Vector Machine (SVM)(vomRadial/Poly)	0.871/0.933	0.905/0.857	0.931/0.903	0.905/0.905	0.80/0.90	0.905/0.905
Nearest Shrunken Centroids (NSC)/vomNSC	0.833/0.90	0.619/0.9524	0.9/0.90	0.905/0.619	0.833/0.833	0.619/0.619

ML classifier (deseq)	1st Method with Vars(100) genes		2nd Method with DEG(100) genes		3rd Method with WGCNA(100) genes	
	Train accuracy	Test accuracy	Train accuracy	Test accuracy	Train accuracy	Test accuracy
pairwise linear discriminant analysis (PLDA)	0.733	0.810	0.6	0.571	0.633	0.857
negative binomial linear discriminant analysis (NBLDA)	0.621	0.762	0.516	0.905	0.700	0.810
vomn-based diagonal linear discriminant analysis (vomDLDA)	0.700	0.810	0.867	0.905	0.700	0.810
vomn-based diagonal quadratic discriminant analysis (vomDQDA)	0.800	0.905	0.837	1	0.833	0.8524