

Resource Summary Report

Generated by [dkNET](#) on Sep 7, 2026

Machine learning approach yields epigenetic biomarkers of food allergy: A novel 13-gene signature to diagnose clinical reactivity

DOI:10.17504/protocols.io.x7pfrmn

Type: Protocol

Proper Citation

Ayush Alag 2019. Machine learning approach yields epigenetic biomarkers of food allergy: A novel 13-gene signature to diagnose clinical reactivity. protocols.io
dx.doi.org/10.17504/protocols.io.x7pfrmn

Protocol Information

URL: <https://dx.doi.org/10.17504/protocols.io.x7pfrmn>

Authors: Ayush Alag

Summary: Current laboratory tests have a less than 50% accuracy in distinguishing between people who have food allergies (FA) and those who are merely sensitized to foods, resulting in the use of expensive and potentially dangerous Oral Food Challenges. Our study presents a purely-computational machine learning approach, conducted using DNA Methylation (DNAm) data, to accurately diagnose food allergies and find genes that are strong biomarkers of the disease. We built two deep learning classifiers with twelve CpG-input features each that achieved perfect accuracy and an AUROC of 1 on the completely hidden cross-validation cohort. In addition, 24 additional classifiers were created that each had an average cross-validation accuracy of 98.35%. These 26 classifiers yielded a total of 18 unique CpGs, which mapped to 13 genes that are strong epigenetic biomarkers of FA. Biological enrichment on the 13-gene signature yielded new insights. Notably, our FA-discriminating genes were strongly associated with the immune system, which helps validate our findings. Seven of the 13 genes overlapped with previous food-allergy and DNAm studies. Previous studies have also created a perfect classifier for this dataset, but they used a 96-CpG input feature set built on both data-driven and a priori biological insights. Our study is an improvement on previous work because it maintains a perfect classification accuracy using only 18 highly discriminating CpGs (0.005% of the total available features). In machine learning, simpler models, as used in our study, are preferred over more complex ones (all other things being equal). In addition, our completely data-driven approach eliminates the need for *a priori* information and allows for generalizability to DNAm classification problems in other disease areas, which may result in novel gene associations or accurate diagnostic tests for those diseases.

Associated Publications: Alag A (2019) Machine learning approach yields epigenetic biomarkers of food allergy: A novel 13-gene signature to diagnose clinical reactivity. PLoS ONE 14(6): e0218253. doi: [10.1371/journal.pone.0218253](https://doi.org/10.1371/journal.pone.0218253)

Affiliations: The Harker School

External URL: <https://doi.org/10.1371/journal.pone.0218253>

Version: 2

Publication Date: 2019

Proper Citation: Ayush Alag 2019. Machine learning approach yields epigenetic biomarkers of food allergy: A novel 13-gene signature to diagnose clinical reactivity. protocols.io dx.doi.org/10.17504/protocols.io.x7pfrmn

Record Creation Time: 20210329T220529+0000

Record Last Update: 20210329T220940+0000

Data and Source Information

Source: