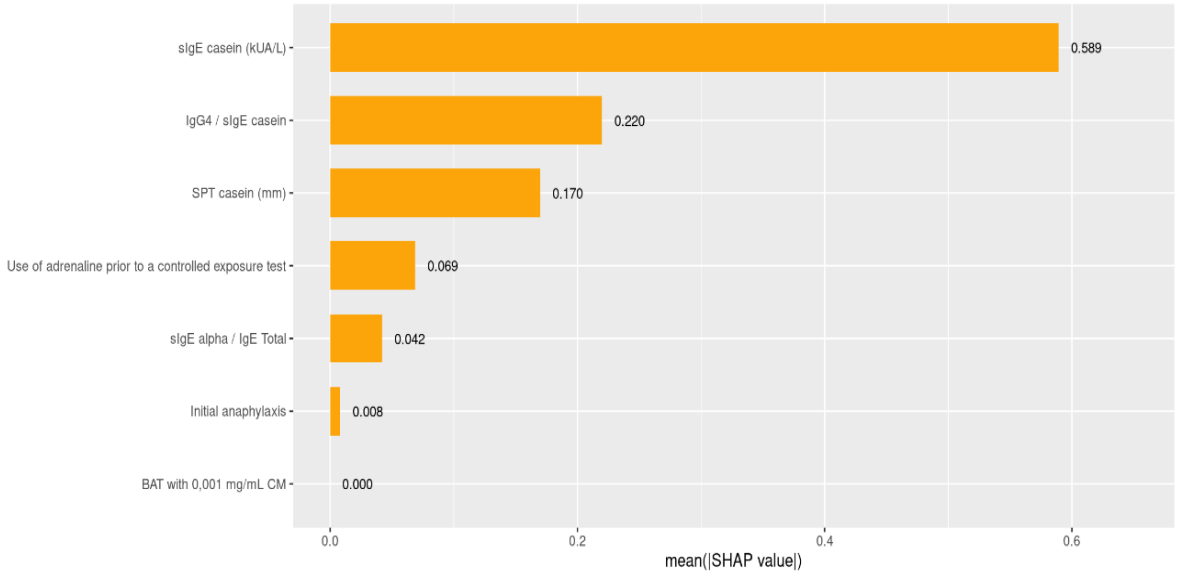
SUPPLEMENTARY MATERIAL

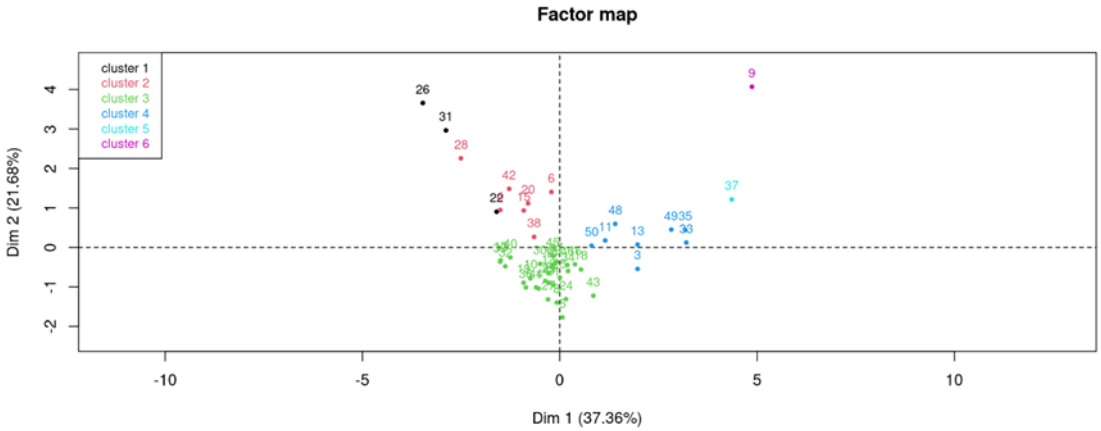
Figure S1. Markers Importance according by XGBoost with Mean absolute SHAP (SHapley Additive exPlanations) values and PCA.



The XGBoost-derived SHAP values indicate the relative importance of markers in predicting BM-tolerant or BM-allergy: a) sIgE casein has the highest mean SHAP value (0.589), suggesting it is the most important predictor; b) IgG4/sIgE casein and SPT casein (mm) are the second and third most important predictors (0.220 and 0.170, respectively); c) BAT with 0.001 mg/mL CM has the lowest mean SHAP value (0.000), indicating minimal predictive importance in this model.

doi: 10.18176/jiaci.1074

Figure S2. Factor map of patient clusters.



This factor map displays the distribution of patients across six distinct clusters. The clustering is based on the same dimensions by Principal Component Analysis (PCA) plot of patients, with tolerance status as a supplementary variable.

Cluster 3 (green) appears to be the largest and most tightly grouped cluster, suggesting a potentially more homogeneous subgroup of patients. Other clusters, such as clusters 1 (black) and 2 (red), show more dispersed patterns, indicating greater variability within these groups. Based on the analysis, cluster 3 shows significant differences from other clusters. The tolerance rate is 73% in cluster 3 compared to 50% in others (p-value 0.092). In the BAT with 0.001 mg/mL CM, 83% of cluster 3 scored 1 versus 45% in others, with a p-value <0.001, indicating higher reactivity. No patient in cluster 3 used adrenaline prior to controlled exposure, unlike 45% in other clusters, with a p-value <0.001. The Alpha-sIgE/ $_{t}$IgE ratio was lower in cluster 3, with a median of 0.00 compared to 0.01 in others (p-value 0.032). These findings highlight distinct immunological background and allergic responses, supported by statistical tests such as Pearson's Chi-squared and Wilcoxon rank-sum.

doi: 10.18176/jiaci.1074