Allergology International 75 (2026) 150–157

Contents lists available at ScienceDirect

Allergology International

journal homepage: http://www.elsevier.com/locate/alit

Original Article

Association of allergen signatures with individualized allergic phenotypes

Dachan Kim [a], Hyung-Ju Cho [a, b], Chang-Hoon Kim [a, b], Min-Seok Rha [a, b, *]

[a] Department of Otorhinolaryngology, Yonsei University College of Medicine, Seoul, South Korea
[b] The Airway Mucus Institute, Yonsei University College of Medicine, Severance Hospital, Seoul, South Korea

ARTICLE INFO

Article history:
Received 20 July 2025
Received in revised form
26 August 2025
Accepted 8 September 2025
Available online 15 October 2025

Keywords:
Allergen
Allergic rhinitis
Asthma
Atopic dermatitis
Immunoglobulin E

Abbreviations:
AR, allergic rhinitis; IgE, immunoglobulin E; HDM, house dust mite; NMF, non-negative matrix factorization; MAST, Multiple Antigen Simultaneous Test; EMR, electronic medical record; ICD-10, International Classification of Diseases, 10th Revision; NNDSVD, non-negative double singular value decomposition

ABSTRACT

Background: Allergen sensitization patterns are heterogeneous, and their clinical relevance is often obscured by extensive cross-reactivity. We applied non-negative matrix factorization (NMF) to disentangle overlapping immunoglobulin E (IgE) signals and define clinically meaningful allergen signatures in a large Korean cohort.
Methods: We analyzed 45,065 patients who underwent multiplex allergen testing (35 inhalants and food components) between 2010 and 2025. Class-scaled specific IgE values (0–6) were factorized by NMF (k = 4). Signature weights were related to asthma, allergic rhinitis, and atopic dermatitis using multivariable logistic regression and to peripheral eosinophil counts and total IgE using age- and sex-adjusted linear models.
Results: Four signatures—mite, grass/weed, pet, and tree—explained 77.7 % of the variance in sensitization. The mite signature predominated (57.6 % of patients) and was strongly associated with allergic rhinitis (adjusted OR: 7.21, 95 % CI: 5.66–9.16), as well as marked increases in eosinophils and total IgE. The pet signature was the strongest predictor of asthma (OR: 8.90, 6.48–12.24). The tree signature showed the strongest association with atopic dermatitis (OR: 6.27, 3.81–10.32) and broader multisystem allergic morbidity. The grass/weed signature exhibited a biphasic age trajectory with a late-adult resurgence but had modest clinical impact. All signatures were significant and graded as determinants of blood eosinophil counts and IgE levels.
Conclusions: Data-driven factorization of multiplex IgE panels yields portable allergen signatures that refine attribution of asthma, allergic rhinitis, and atopic dermatitis and link serologic patterns to systemic inflammation.
© 2025 Japanese Society of Allergology. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Allergic diseases such as asthma, allergic rhinitis (AR), and atopic dermatitis are highly prevalent chronic conditions driven by abnormal immune responses.[1–4] A central feature of these atopic disorders is allergen sensitization—the production of specific immunoglobulin E (IgE) antibodies against environmental and food allergens—which is widely recognized as a key risk factor in their pathophysiology.[5–7] Numerous epidemiological studies have documented strong associations between IgE sensitization and the presence of asthma or AR.[2,5] However, the strength of these associations varies significantly among individuals. To date, no reliable IgE-based biomarker can predict disease development or stratify patients according to future risk.[8] In other words, despite the clear importance of allergen sensitization, clinicians still lack a practical serological indicator to predict whether a sensitized patient will develop asthma, hay fever, or eczema or remain asymptomatic.

The relationship between allergens that elicit sensitization and the subsequent disease phenotype is complex but crucial. Allergic sensitization patterns can influence the clinical manifestations and severity of allergic diseases. For example, polysensitization (IgE reactivity to more than one allergen) is frequently observed and is associated with early onset of allergic diseases, severe symptoms, and increased multimorbidity.[9–12] Sensitization to house dust mite (HDM) has been linked to an increased risk of asthma,[2,13] whereas soybean or milk positivity has been associated with AR multimorbidity.[12] However, systematic research on major allergen sensitization patterns and their impact on allergic diseases in the general population remains limited. The interpretation of

* Corresponding author. Department of Otorhinolaryngology, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea.
E-mail address: msrha@yuhs.ac (M.-S. Rha).

https://doi.org/10.1016/j.alit.2025.09.003
1323-8930/© 2025 Japanese Society of Allergology. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

component-level data is further complicated by extensive cross-reactivity among homologous allergen proteins and demographic or exposure-related confounding factors that can mask true biological associations. Although several large-scale studies in East Asia, including Korea[5,14,15] and China,[16] have profiled sensitization prevalence, their conventional clustering approaches struggled to parse these intertwined effects, and therefore rarely yielded clinically actionable links to specific outcomes.

To overcome these challenges, we treated the high-dimensional IgE matrix as a latent mixture problem and applied non-negative matrix factorization (NMF), an unsupervised deconvolution technique that separates overlapping IgE signals into additive orthogonal components. Using this strategy, we performed an unbiased analysis of allergen sensitization patterns in a large cohort comprising 45,065 Korean patients. With a comprehensive panel of more than 35 food and inhalant allergens, we applied NMF to identify the latent allergen signatures in this population. We then examined how these distinct signatures were related to the prevalence of asthma, AR, and atopic dermatitis. By elucidating the intricate relationship between molecular sensitization patterns and specific disease prevalence, this study aimed to advance our understanding of the pathophysiology of allergic diseases and support the development of more precise diagnostic and prognostic strategies.

Methods

Study population and data acquisition

This retrospective cross-sectional analysis was conducted at Yonsei University Severance Hospital, a tertiary academic medical center in Seoul, Korea. The laboratory database was queried for all AdvanSure™ Allostation Multiple Antigen Simultaneous Test (MAST; LG Life Science, Korea) assays performed between January 2010 and June 2025, yielding 105,962 test panel results. We included all individuals who underwent the MAST panel irrespective of prior allergic disease diagnoses, thereby capturing both symptomatic and screening/non-allergic cases to minimize selection bias. To establish a uniform baseline, only the earliest panel for each individual was retained, reducing the dataset to 90,611 unique patients. The assay originally quantified serum IgE against 123 inhalant or food allergens; 87 allergens with missing results in ≥40 % of panels were excluded, leaving 35 allergens for analysis. Panels with missing IgE values for the retained allergens were excluded, resulting in a final analytic cohort of 45,065 patients. In this final cohort, the proportions with ICD-10-coded allergic diseases were 68.8 % for AR, 29.9 % for asthma, and 5.4 % for atopic dermatitis (Table 1). The study complied with the Declaration of Helsinki and was approved by the Institutional Review Board of Yonsei University College of Medicine (IRB 4-2025-0708). The requirement for informed consent was waived owing to the retrospective nature of the study, and all data were anonymized through IRB-approved procedures. Demographic characteristics and baseline laboratory findings of the study subjects are presented in Table 1.

Allergen-specific IgE quantification

The 35 retained allergens were classified into seven groups: HDM, tree pollen, grass pollen, weed pollen, animal dander, fungi, and insects. Indoor allergens included HDM (*Dermatophagoides farinae* [Der f] and *Dermatophagoides pteronyssinus* [Der p]), animal dander (cat and dog), and insects (cockroach mix). Outdoor allergens included tree pollen (acacia, ash mix, birch alder mix, hazelnut, Japanese cedar, oak, pine, poplar mix, sallow willow, and sycamore mix), grass pollen (Bermuda grass, orchard grass, reeds, rye grass, sweet vernal grass, and timothy grass), weed pollen (dandelion, goldenrod, Japanese hop, mugwort, oxeye daisy, ragweed, Russian thistle, and pigweed), and fungi (*Alternaria*, *Aspergillus*, *Cladosporium*, and *Penicillium*). The test results were classified into seven levels: class 0 (0.00–0.34 kU/L), class 1 (0.35–0.69 kU/L), class 2 (0.70–3.49 kU/L), class 3 (3.50–17.49 kU/L), class 4 (17.50–49.99 kU/L), class 5 (50.00–99.99 kU/L), and class 6 (>100 kU/L).

Table 1
Demographic and clinical characteristics of the study participants.

	Overall
Total number	45,065
Sex	Male: 24,975 (55.4 %)
	Female: 20,087 (44.6 %)
Age	39.8 (39.6–40.0)
Cluster	Nonallergic: 18,686 (41.5 %)
	Mite: 15,205 (33.7 %)
	Pet: 4804 (10.7 %)
	Grass: 3977 (8.8 %)
	Tree: 2393 (5.3 %)
Asthma	13,453 (29.9 %)
Allergic rhinitis	31,025 (68.8 %)
Atopic dermatitis	2452 (5.4 %)
Total IgE mean (95 % CI)	250.2 (244.3–256.1)
Eosinophil count (95 % CI)	227.6 (224.1–231.1)
Eosinophil % (95 % CI)	3.2 (3.2–3.2)

IgE, immunoglobulin E; CI, confidence interval.

Clinical laboratory variables

Blood eosinophil percentage, absolute eosinophil count, and total serum IgE levels were retrieved from the institutional electronic medical record (EMR). To minimize temporal heterogeneity, only the first laboratory measurement obtained for each patient, defined as the assay performed on or closest to the date of the index MAST, was retained for analysis.

Clinical diagnoses of allergic diseases

As described in previous studies,[17,18] asthma (J45–46), AR (J30.1–30.4), and atopic dermatitis (L20) were identified from the EMR using International Classification of Diseases, 10th Revision (ICD-10) billing codes. A patient was considered positive for a given condition if at least one relevant code had ever been assigned at an outpatient or inpatient encounter. This one-code-ever approach maximized sensitivity while minimizing misclassification in such a large retrospective cohort.

Co-sensitization network construction

To elucidate population-level co-sensitization patterns, the proportion of participants who were simultaneously sensitized to both allergens was first computed for every allergen pair; self-pairs were set to zero. Edges with a co-sensitization proportion below 0.05 were pruned, and the retained values were used as edge weights in an undirected graph built with NetworkX (Python 3.10). Each node was annotated with its prevalence (which later drove the node size) and a manually curated allergen class (mite, pet, insect, mold, tree, grass, weed, or food) that determined the node color. The resulting GraphML file, together with matching node and edge tables, was imported into Cytoscape 3.10 and visualized using the edge-weighted spring-embedded layout, which positions highly co-sensitized allergens close together while respecting edge weights.

NMF and visualization of allergen signatures

After excluding 18,686 patients whose serum IgE (sIgE) profiles were completely negative across all 35 allergens, the analytical matrix comprised 26,379 patients × 35 allergens. Class-scale sIgE values (0–6) obtained from the MAST assay were used in their original form. NMF was performed with the sklearn implementation (version 1.3.0) using non-negative double singular value decomposition (NNDSVD) initialization and a maximum of 2000 iterations to ensure convergence. Because reconstruction error decreased monotonically with k and no distinct "elbow" was observed, we prioritized biological interpretability and parsimony; $k = 4$ was selected as it yielded stable, non-redundant, and clinically coherent components across random initializations. The factorization yielded a sample × component matrix W (patient signature scores) and a component × allergen matrix H (signature loadings). The dominant signature for each allergen was defined as the component with the highest loading, and this labeling was propagated to subsequent clinical association analyses.

Results

Class-based co-sensitization network

The allergen co-sensitization network graph comprised 35 allergen nodes connected by 357 weighted edges, giving an overall network density of 0.60, meaning that 60 % of all possible allergen pairs shared measurable co-sensitization. Spatially, the edge-weighted spring-embedded layout placed biologically related allergens into tight clusters (Fig. 1). Nodes belonging to the mite class formed a dominant central hub and exhibited the largest diameters, confirming that HDM sensitization was the most common and interconnected allergen group in this population. Quantitatively, *D. farinae* had the highest node prevalence (1.23 on a class 0–6 scale), whereas Mackerel was the least common allergen (0.006). The strongest pairwise co-sensitization occurred between *D. farinae* and *D. pteronyssinus* (edge weight ≈ 3.98), whereas the weakest retained edge above the 0.05 threshold linked hop and soybean (edge weight ≈ 0.05). Grass, weeds, pets, trees, mold, food, and insect allergens were organized into distinct peripheral subnetworks surrounding this core. These class-based islands were linked to the mite hub by thinner inter-class edges, reflecting weaker co-sensitization frequencies. Therefore, the network visualizes a hierarchy in which HDM allergens act as major hubs, while other classes maintain more localized, within-class co-sensitization patterns.

Data-driven extraction of allergen signatures

Exploratory NMF of raw sIgE titers identified four reproducible allergen signatures that together explained 77.7 % of the variance in the sensitization matrix (Frobenius reconstruction error = 465; Fig. 2). Signature composition closely mirrored biological groupings, allowing for an intuitive nomenclature. Specifically, the first component, dominated by the two HDM species and tropomyosin-bearing indoor allergens, was designated as the "mite signature." The second component, driven almost exclusively by grass (rye, timothy, orchard, and Bermuda pollens) and weed (ragweed and pigweed) pollens, was termed the "grass/weed signature." The third component was characterized by appreciable input from cat

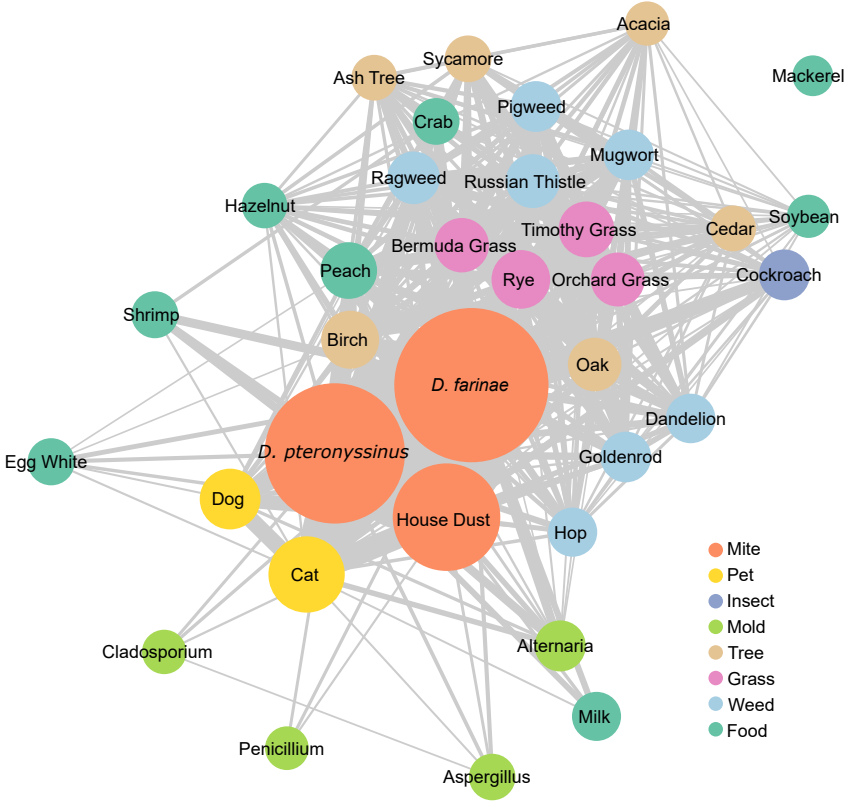

Fig. 1. Allergen co-sensitization network. Each node represents an individual allergen and is colored by allergen class (mite, pet, insect, mold, tree, grass, weed, or food). Node size is proportional to the population prevalence of specific-IgE positivity for that allergen. Edge thickness denotes the strength of pairwise correlation between specific-IgE levels; thicker edges indicate stronger correlations, highlighting common co-sensitization patterns.

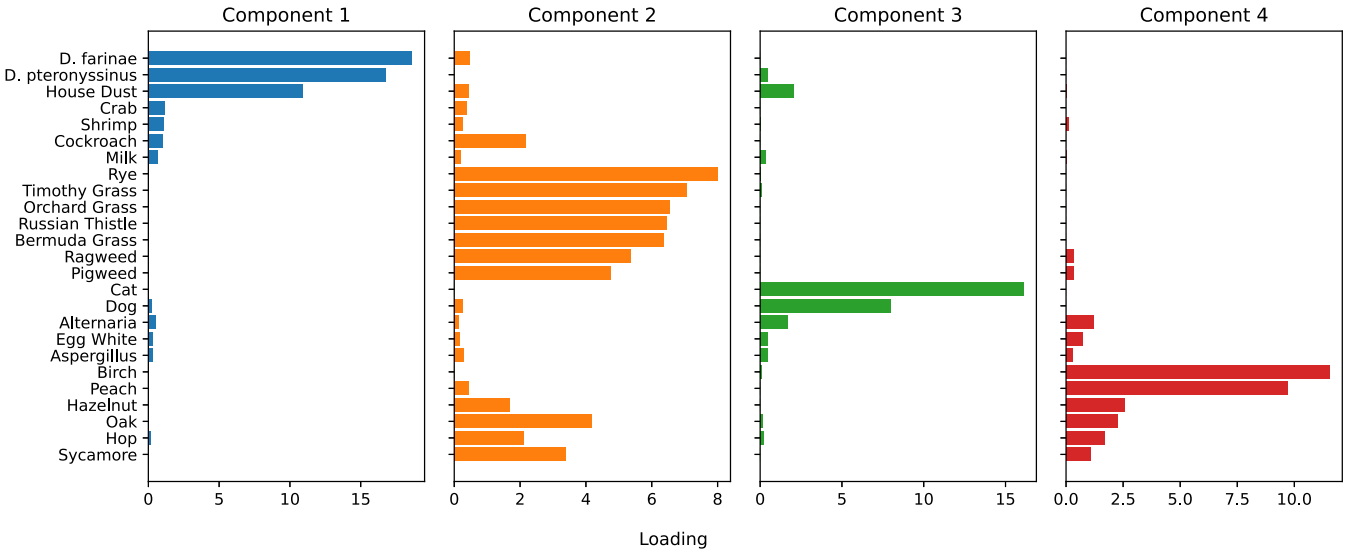

Fig. 2. Top allergen loadings for each non-negative matrix factorization signature. Horizontal bar charts display the allergens contributing most to each component (comp 1–4). Bar length equals loading coefficient magnitude; longer bars indicate greater influence. Bars are colored by component (blue, orange, green, red) and ordered from strongest to weakest, emphasizing interpretable allergen composition of every signature.

and dog dander, and was therefore labeled the "pet signature." Finally, the fourth component centered on tree pollens such as oak and birch and was named the "tree signature." These four signatures are readily interpretable and mirror the major hubs observed in the population-level co-sensitization network (Fig. 1). The full 35 × 4 loading matrix (H table) is provided in Supplementary Table S1 to facilitate rapid application of this framework to independent cohorts or single-patient analyses.

Distribution of allergen signatures across patients

Heatmap visualization of individual signature weights (Fig. 3) showed that most patients were driven by a single dominant signature, although only 3.4 % met the definition of a mixed profile (≥0.25 loading in at least two signatures). The mite signature (comp 1) predominated, accounting for 57.6 % of patients, with the grass/weed, pet, and tree signature were observed in 15.1 %, 18.2 %, and 9.1 % of patients, respectively. Compared with the single-signature group, the mixed-signature group exhibited significantly elevated immunologic activity, as indicated by higher total IgE levels (mean 1059 vs 340 kU/L) and blood eosinophil counts (355 vs 256 cells/μL; both $p < 0.001$; Supplementary Fig. S1), alongside a greater prevalence of allergic rhinitis (78.4 % vs 71.8 %, $p < 0.001$) and atopic dermatitis (16.6 % vs 6.2 %, $p < 0.001$), while asthma prevalence remained statistically comparable (30.9 % vs 28.1 %; not significant). These patterns align with prior reports that allergen polysensitization correlates with increased symptom burden and disease severity.[33]

Age-dependent dynamics of allergen signatures

Given that allergic phenotypes and sensitization are known to be influenced by age,[14,19,20] we investigated age-related dynamics of allergic sensitization patterns. LOWESS-smoothed curves (Fig. 4) revealed signature-specific age trajectories with well-defined peak ages. The mite signature reached its maximum average contribution at 14 years of age and declined progressively across adulthood. Notably, the grass/weed signature displayed a biphasic pattern: an initial peak at 11 years of age, a trough in early adulthood, and a secondary increase peaking at 56 years of age, suggesting the re-emergence of grass or weed sensitization in later life. The pet signature peaked slightly later, at 16 years of age, before tapering off with age. The tree signature exhibited a modest early peak at 11 years of age and remained low but detectable throughout the lifespan. These discrete peak ages may highlight the developmental windows during which each allergen signature exerts its greatest immunological influence.

Immunologic correlates of allergen signatures

Because allergic diseases are driven by IgE-mediated hypersensitivity and are closely associated with eosinophilic inflammation,[21–23] we further analyzed the relationship between allergic signatures and blood eosinophil counts and IgE levels. Age- and sex-adjusted linear models quantified the impact of each signature on blood eosinophil count (cells per μL) (Fig. 5A). Per unit increase in signature weight, the mite signature increased eosinophil count by β = 406 (95 % CI, 367–445), closely matched by the pet signature (β = 432, 381–483). The tree signature showed an intermediate effect (β = 259, 208–311), whereas the grass/weed signature exerted the smallest, but still significant, increment (β = 149, 109–189). All associations were statistically significant ($p < 0.001$). For total IgE (kU/L), the effect sizes were similarly substantial but displayed a slightly different ranking (Fig. 5B). The mite signature yielded the largest increment (β = 2367 kU/L, 2302–2432), followed by the pet (β = 1517 kU/L, 1438–1597), the grass/weed (β = 1375 kU/L, 1308–1441), and tree (β = 1282 kU/L, 1206–1358) signatures. All β coefficients were statistically significant ($p < 0.001$).

Association of allergen signatures with disease entity

Finally, we investigated the association between allergen signatures and each allergic disease. Multivariable logistic regression models (Fig. 6) revealed signature-specific disease patterns. For asthma, the pet signature showed the greatest association (OR = 8.90; 95 % CI, 6.48–12.23; $p < 0.001$), followed by the tree signature (OR = 3.43, 2.41–4.89, $p < 0.001$). In contrast, the mite

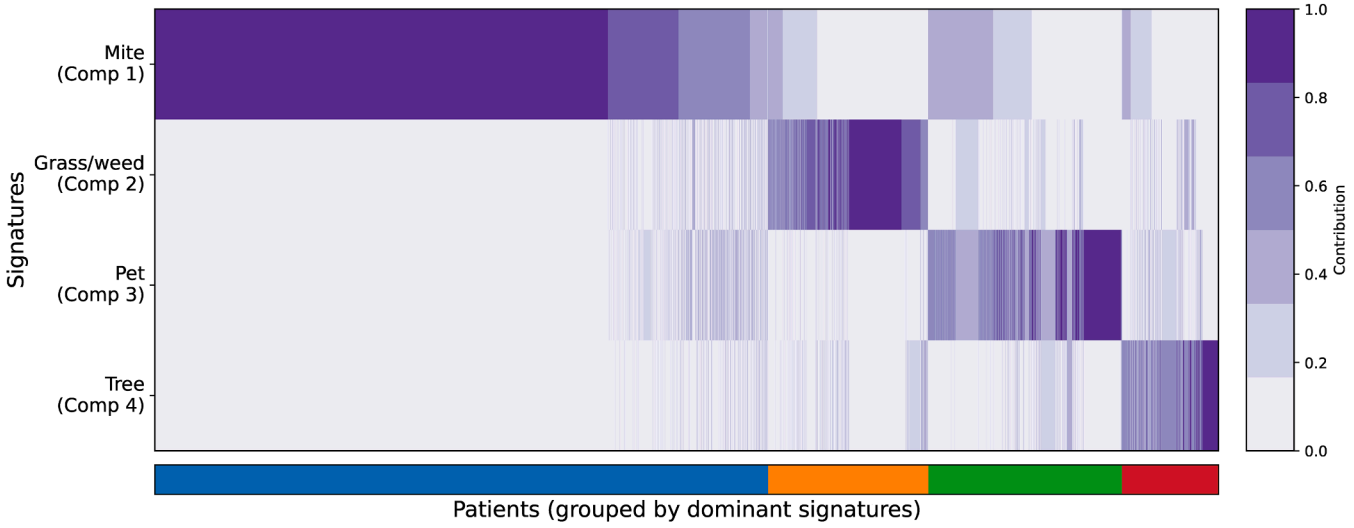

Fig. 3. Heatmap of patient-level allergen signature weights. Columns represent individual patients, ordered by their dominant signature (color strip below: blue = comp 1, orange = comp 2, green = comp 3, red = comp 4). Rows correspond to the four signatures. Cell color intensity (0–1) indicates the normalized contribution of each signature to a patient's overall IgE profile.

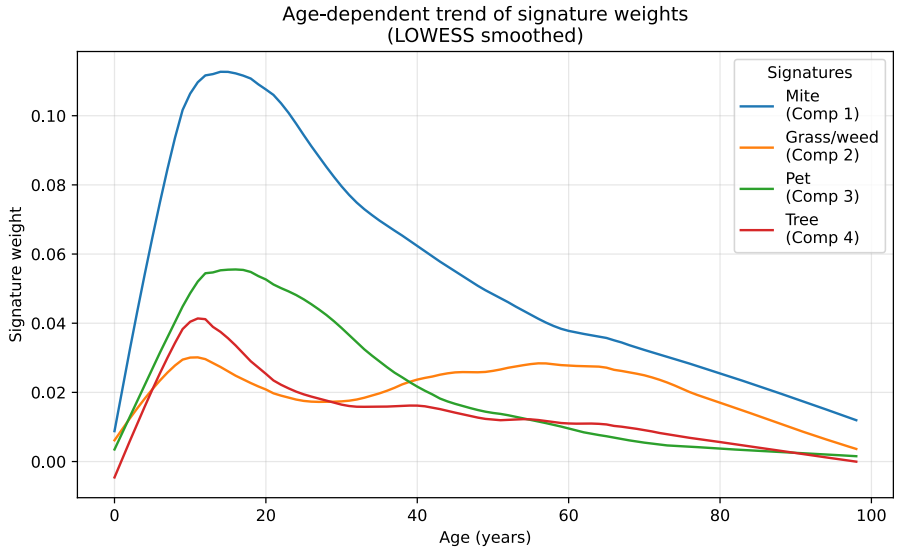

Fig. 4. Trends in allergen signature weights according to age. LOWESS-smoothed lines depict mean signature contributions across age (0–100 years). Colors match signature IDs. The mite signature peaks in childhood and declines with age, whereas the grass (comp 2) and pet (comp 3) signatures rise in adolescence, and the tree signature (comp 4) shows a modest late-life increase.

and grass/weed signatures were not significantly associated with asthma (OR = 0.83, 0.64–1.07 and OR = 0.99, 0.76–1.29, respectively). In AR, the mite signature exhibited the strongest association (OR = 7.21, 5.66–9.17, $p < 0.001$), while the pet and tree signatures had moderate associations (OR = 3.03, 2.17–4.25 and OR = 3.59, 2.44–5.29). The grass/weed signature was not significantly associated with AR (OR = 0.99, 0.75–1.30). For atopic dermatitis, the tree signature was most strongly associated with the disease (OR = 6.27, 3.81–10.33, $p < 0.001$), while the mite (OR = 4.98, 3.33–7.45), pet (OR = 4.14, 2.56–6.70), and grass/weed (OR = 1.61, 1.04–2.49) signatures were associated with smaller but still significant risks. Altogether, these findings suggest that pet and tree sensitization profiles are related to multi-organ allergic morbidity, mite sensitization is strongly associated with AR, and grass/weed sensitization exhibits relatively modest associations with allergic diseases. We further analyzed the association between each allergen signature and allergic multimorbidity (Supplementary Fig. S2). Pet sensitization conferred the highest multimorbid risk, with an OR of 9.22 (95 % CI, 6.57–12.94) for asthma + rhinitis and 12.91 (5.68–29.31) for the triple-disease combination. Tree sensitization also showed robust effects with an OR of 9.31 (4.29–20.20) for asthma + dermatitis and 9.84 (4.28–22.63) for all three diseases. In contrast, mite and grass/weed signatures were comparatively less associated with multimorbidity, with maximum ORs of 6.84 and 2.25, respectively.

Discussion

Our study demonstrates that NMF can disentangle overlapping IgE signals in a large cohort of more than 45,000 individuals and reveals four biologically coherent allergen signatures—mite, grass/weed, pet, and tree—that explain nearly 80 % of the variance in a

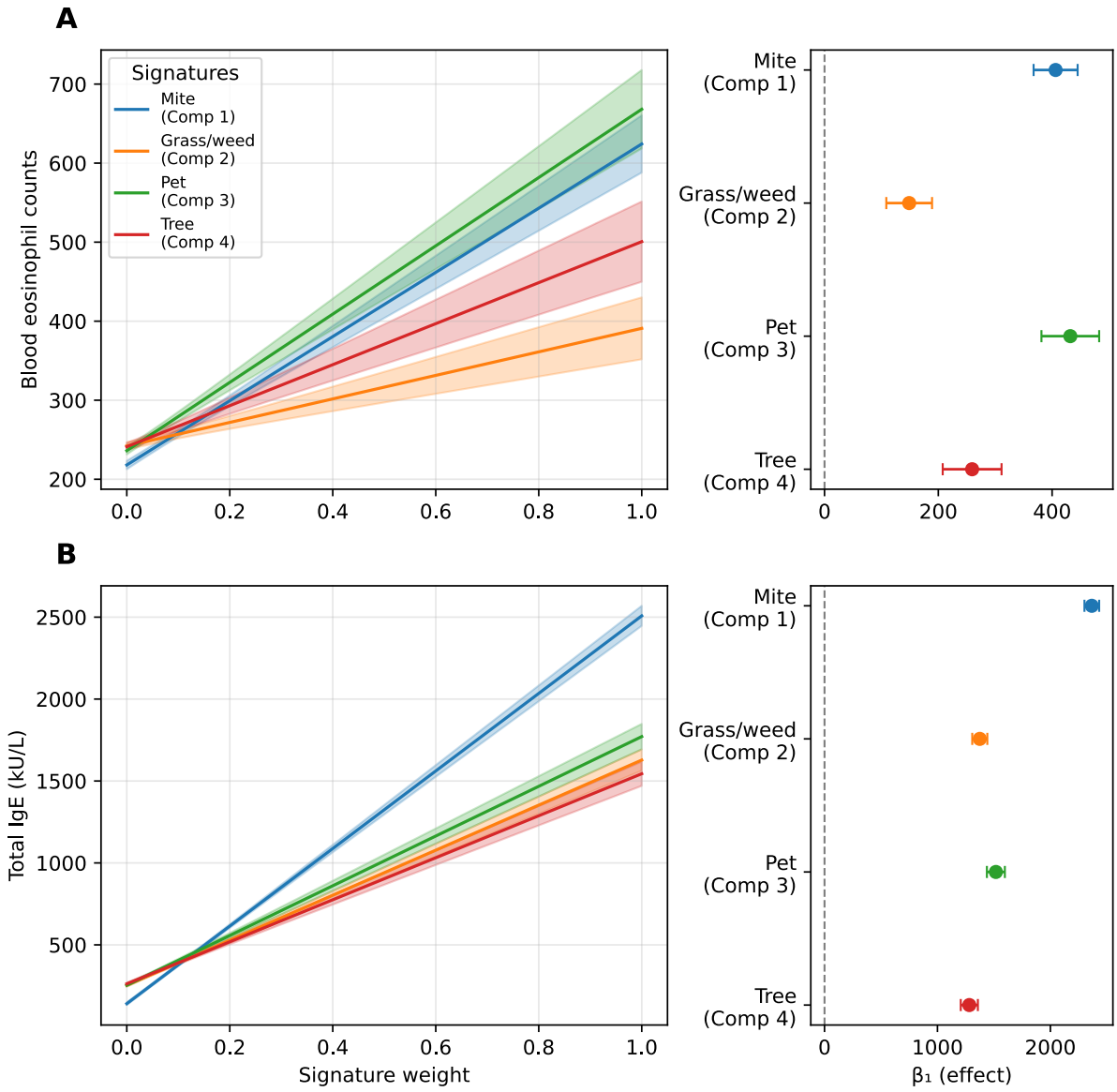

Fig. 5. Associations between allergen signatures and blood eosinophil counts and IgE levels. **(A)** Signature weights vs. blood eosinophil counts. **(B)** Signature weights vs. total serum IgE. Left: LOWESS-smoothed regression lines with 95 % confidence ribbons. Right: Multivariable β coefficients (circles) and 95 % CIs (horizontal bars) from age- and sex-adjusted linear models, representing the expected change in the outcome per 1-unit increase in signature weight. Vertical dashed line marks no association; colors denote signatures.

35-component sensitization panel. By deconvoluting cross-reactive patterns that obscured previous analyses, we assigned each patient a high-resolution serologic "allergen fingerprint" and linked these fingerprints to distinct clinical and immunologic phenotypes. This represents a substantive advance over earlier clustering studies that grouped patients by raw positivity profiles without correcting for cross-reactivity or confounders and therefore reported less specific, often population-dependent patterns.

The allergen signatures identified in this study have important clinical implications. The mite signature demonstrated the strongest association with AR and marked peripheral eosinophilia/IgE elevations; however, it showed no independent association with asthma after adjusting for age and sex. This indicates that earlier univariable studies may have over-attributed asthma risk to mite sensitization.[10,13] However, in our adult cohort with near-ubiquitous HDM exposure and extensive polysensitization, asthma susceptibility is better explained by the pet and tree signatures. Conversely, the pet signature emerged as the most powerful predictor of asthma. Pet dander allergens are exceptionally small and remain airborne for prolonged periods, allowing major cat (Fel d 1) and dog (Can f 1, Can f 5) proteins to reach the bronchial mucosa, unlike larger pollen or mite particles that are typically deposited in the upper airways.[24,25] Once in the lower airway, these proteins act as powerful Th2-skewing adjuvants, driving IgE class switching, eosinophil recruitment, and airway hyper-responsiveness.[26,27] The tree signature, enriched in pathogenesis-related class 10 proteins—Bet v 1 (birch) and Cor a 1 (hazelnut)—and plant lipid-transfer proteins—Pru p 3 (peach)—captures IgE cross-reactivity across pollen and botanically related foods. These homologous proteins constitute the molecular backbone of pollen–food syndrome, and IgE primed by inhaled tree pollen recognizes structurally similar epitopes in fruits, nuts, and vegetables, eliciting oral or systemic reactions upon ingestion.[28–30] This airway-to-gut-to-skin axis explains why high

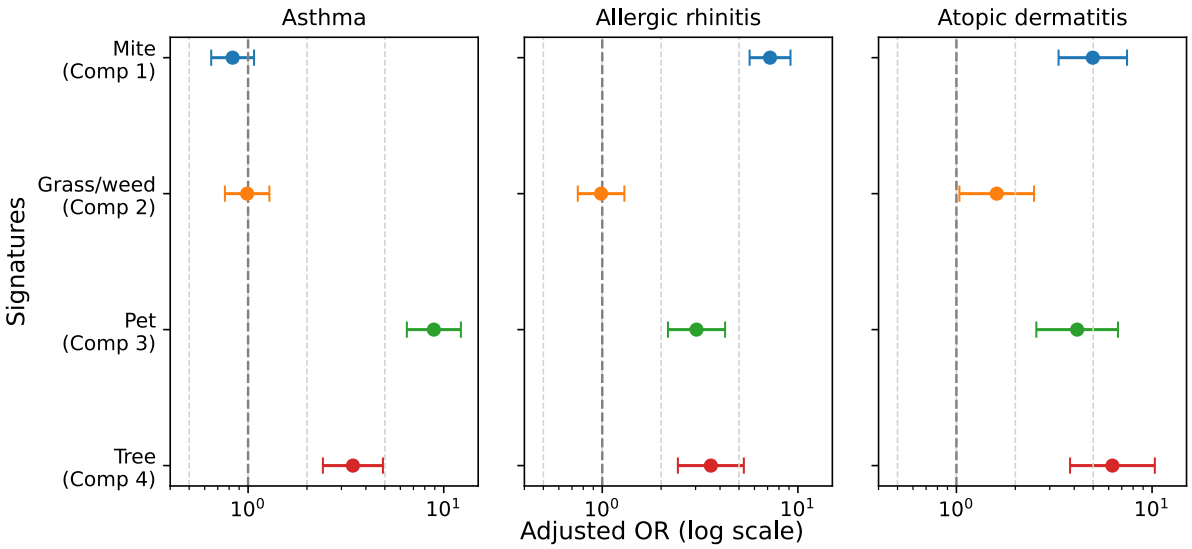

Fig. 6. Adjusted odds ratios for allergic diseases. Forest plots display age- and sex-adjusted odds ratios (log scale) with 95 % CIs for asthma, allergic rhinitis, and atopic dermatitis. Each row represents a non-negative matrix factorization signature (colors as above). The vertical dashed line (OR = 1) denotes no association; points to the right (OR > 1) indicate increased odds, and points to the left (OR < 1) indicate decreased odds with higher signature weight.

tree signature weights in our cohort strongly predict atopic dermatitis and broader multisystem allergic morbidity. The grass/weed signature showed a modest association with atopic dermatitis but was not significantly associated with asthma or AR. Furthermore, it had the weakest impact on eosinophilia among all signatures. Although a considerable portion of patients (15.1 %) exhibited the grass/weed signature, its overall contribution to allergic diseases may be relatively limited in Korean patients.

Previous studies have consistently reported that the prevalence of allergic sensitization decreases with increasing age.[2,14,20] Similarly, we found that the peak age of all allergen signatures was approximately 10–20 years old, followed by a gradual decrease. Intriguingly, the grass/weed signature displayed a biphasic age trajectory with a late-life resurgence. This result may derive from a potential increase in outdoor activities, such as gardening after retirement, coupled with age-related alterations in the immune system, including diminished type-2 immune regulation and regulatory T-cell efficacy,[31,32] thereby lowering the threshold for novel sensitization—a phenomenon infrequently documented in previous research.

Methodologically, our use of class-level MAST data in the native 0–6 scale avoids the assumptions imposed by arbitrary log transformations, while NNDSVD initialization enhances the reproducibility of the NMF solution. Integrating prevalence-weighted network analysis with signature-specific regression further bridges population-level patterns and individual-level risk, providing a template for precision allergy diagnostics. We provide a 35 × 4 loading matrix (H table) for the four signatures (Supplementary Table S1). This will allow external investigators to estimate signature weights in new cohorts, or even in a single patient, by simple matrix multiplication of their class 0–6 IgE vector using this public H table, without having to rerun NMF. The portability of the H table enables rapid external validation and facilitates its application in prospective clinics, biobanks, and geographically distinct allergen panels.

This study had several limitations. First, the single-center, retrospective design and reliance on ICD-10 billing codes for diagnosis may have led to misclassification of disease status despite our one-code-ever rule. Although this potential misclassification is unlikely to be systematically related to the NMF-derived signature structure, it may slightly affect the observed associations between allergen signatures and each disease entity. Second, the proportion of patients varied across allergic disease entities, which may have influenced the results. Third, this study lacked longitudinal IgE data from individual patients to test the temporal stability of the signatures. Future prospective cohorts with repeated MAST testing are required to quantify the within-person stability of NMF-derived signature weights and to evaluate their predictive value for disease onset, exacerbations, and progression. Incorporating in vivo functional readouts (e.g., basophil activation and skin prick tests) will help clarify causal pathways and guide selection and monitoring of allergen-specific immunotherapy.

In summary, decoding high-dimensional IgE landscapes with NMF reveals individualized allergen signatures that capture immunologic heterogeneity, predict organ-specific disease risks, and lay the groundwork for precision management of allergic disorders.

Acknowledgements

This study was supported by the Bio & Medical Technology Development Program of the National Research Foundation of Korea (NRF), funded by the Korea government (MSIT) (RS-2024-00439160), and by the NRF grant funded by the Korea government (MSIT) (RS-2025-00516950).

MID (Medical Illustration & Design), as a member of the Medical Research Support Services of Yonsei University College of Medicine, providing excellent support with medical illustration.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.alit.2025.09.003.

Conflict of interest

The authors have no conflict of interest to declare.

Authors' contributions

DK: Study design, Data collection, Data analysis, Writing original draft, Reviewing, Revising the manuscript.

HJC: Study design, Data collection, Reviewing, Revising the manuscript.

CHK: Study design, Data collection, Reviewing, Revising the manuscript.

MSR: Study design, Data collection, Data analysis, Funding acquisition, Reviewing, Revising the manuscript.

References

1. Wise SK, Damask C, Roland LT, Ebert C, Levy JM, Lin S, et al. International consensus statement on allergy and rhinology: Allergic rhinitis – 2023. *Int Forum Allergy Rhinol* 2023;**13**:293–859.
2. Chang SW, Lee HY, Kim JH, Chang JH, Choi HS, Kang JW. Sensitization of common allergens and co-sensitization patterns in the Korean population. *Sci Rep* 2025;**15**:13664.
3. Foong R-X, du Toit G, Fox AT. Asthma, food allergy, and how they relate to each other. *Front Pediatr* 2017;**5**:89.
4. Mersha TB, Afanador Y, Johansson E, Proper SP, Bernstein JA, Rothenberg ME, et al. Resolving clinical phenotypes into endotypes in allergy: Molecular and omics approaches. *Clin Rev Allergy Immunol* 2021;**60**:200–19.
5. Yun J-E, Ko EB, Jung HI, Gu K-M, Kim TW, Park S-Y, et al. Allergen sensitization and its association with allergic diseases in the Korean population: Results From the 2019 Korea National Health and Nutrition examination survey. *Allergy Asthma Immunol Res* 2024;**16**:534–45.
6. Zhang W, Xie B, Liu M, Wang Y. Associations between sensitisation to allergens and allergic diseases: a hospital-based case–control study in China. *BMJ Open* 2022;**12**:e050047.
7. Falcon RMG, Caoili SEC. Immunologic, genetic, and ecological interplay of factors involved in allergic diseases. *Front Allergy* 2023;**4**:1215616.
8. Custovic A, Custovic D, Fontanella S. Understanding the heterogeneity of childhood allergic sensitization and its relationship with asthma. *Curr Opin Allergy Clin Immunol* 2024;**24**:79.
9. Bousquet J, Anto JM, Bachert C, Baiardini I, Bosnic-Anticevich S, Walter Canonica G, et al. Allergic rhinitis. *Nat Rev Dis Primer* 2020;**6**:1–17.
10. Burte E, Bousquet J, Siroux V, Just J, Jacquemin B, Nadif R. The sensitization pattern differs according to rhinitis and asthma multimorbidity in adults: the EGEA study. *Clin Exp Allergy J Br Soc Allergy Clin Immunol* 2017;**47**:520–9.
11. González-Pérez R, Poza-Guedes P, Pineda F, Galán T, Mederos-Luis E, Abel-Fernández E, et al. Molecular mapping of allergen exposome among different atopic phenotypes. *Int J Mol Sci* 2023;**24**:10467.
12. Li Y-T, Ye Q-Q, Lu Y-X, Yang K-X, Zhang P-P, Chen C, et al. Allergen sensitization patterns: Allergic rhinitis with multimorbidity versus alone—A real-world study. *Clin Transl Allergy* 2025;**15**:e70030.
13. Valero A, Quirce S, Dávila I, Delgado J, Domínguez-Ortega J. Allergic respiratory disease: Different allergens, different symptoms. *Allergy* 2017;**72**:1306–16.
14. Choi JH, Suh JD, Kim I, Kim JK, Cho JH. Changes in sensitization rates for respiratory and food allergens by age. *Clin Exp Allergy* 2024;**54**:219–21.
15. Park SC, Hwang CS, Chung HJ, Purev M, Al Sharhan SS, Cho H-J, et al. Geographic and demographic variations of inhalant allergen sensitization in Koreans and non-Koreans. *Allergol Int Off J Jpn Soc Allergol* 2019;**68**:68–76.
16. Luo W, Wang D, Zhang T, Zheng P, Leng D, Li L, et al. Prevalence patterns of allergen sensitization by region, gender, age, and season among patients with allergic symptoms in mainland China: A four-year multicenter study. *Allergy* 2021;**76**:589–93.
17. Oh J, Lee M, Kim M, Kim HJ, Lee SW, Rhee SY, et al. Incident allergic diseases in post-COVID-19 condition: multinational cohort studies from South Korea, Japan and the UK. *Nat Commun* 2024;**15**:2830.
18. Oh J, Kim S, Kim MS, Abate YH, Abd ElHafeez S, Abdelkader A, et al. Global, regional, and national burden of asthma and atopic dermatitis, 1990–2021, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. *Lancet Respir Med* 2025;**13**:425–46.
19. Warm K, Backman H, Lindberg A, Lundbäck B, Rönmark E. Low incidence and high remission of allergic sensitization among adults. *J Allergy Clin Immunol* 2012;**129**:136–42.
20. Warm K, Hedman L, Lindberg A, Lötvall J, Lundbäck B, Rönmark E. Allergic sensitization is age-dependently associated with rhinitis, but less so with asthma. *J Allergy Clin Immunol* 2015;**136**:1559–65. e2.
21. Asano K, Ueki S, Tamari M, Imoto Y, Fujieda S, Taniguchi M. Adult-onset eosinophilic airway diseases. *Allergy* 2020;**75**:3087–99.
22. Shamji MH, Valenta R, Jardetzky T, Verhasselt V, Durham SR, Würtzen PA, et al. The role of allergen-specific IgE, IgG and IgA in allergic disease. *Allergy* 2021;**76**:3627–41.
23. Smith KA, Gill AS, Pollard CE, Sumsion JS, Saffari H, Ashby S, et al. An eosinophil peroxidase activity assay accurately predicts eosinophilic chronic rhinosinusitis. *J Allergy Clin Immunol* 2023;**152**:400–7.
24. Torres-Borrego J, Sánchez-Solís M. Dissecting airborne allergens. *J Clin Med* 2023;**12**:5856.
25. Gauvreau GM, El-Gammal AI, O'Byrne PM. Allergen-induced airway responses. *Eur Respir J* 2015;**46**:819–31.
26. Stark JM, Liu J, Tibbitt CA, Christian M, Ma J, Wintersand A, et al. hyper-responsiveness in a model of asthma marked by vigorous TH 2 and TH 17 cell responses. *Allergy* 2022;**77**:2987–3001.
27. Tsolakis N, Malinovschi A, Nordvall L, Mattsson L, Lidholm J, Pedroletti C, et al. Sensitization to minor cat allergen components is associated with type-2 biomarkers in young asthmatics. *Clin Exp Allergy* 2018;**48**:1186–94.
28. Heratizadeh A. Atopic dermatitis: new evidence on the role of allergic inflammation. *Curr Opin Allergy Clin Immunol* 2016;**16**:458.
29. Hoffmann-Sommergruber K, Roesner LM. The clinical impact of cross-reactions between allergens on allergic skin diseases. *Curr Opin Allergy Clin Immunol* 2020;**20**:374.
30. Mittermann I, Wikberg G, Johansson C, Lupinek C, Lundeberg L, Crameri R, et al. IgE Sensitization Profiles Differ between Adult Patients with Severe and Moderate Atopic Dermatitis. *PLOS ONE* 2016;**11**:e0156077.
31. Jagger A, Shimojima Y, Goronzy JJ, Weyand CM. Regulatory T cells and the immune aging process: a mini-review. *Gerontology* 2013;**60**:130–7.
32. Erin Diane Lewis. Age-associated alterations in immune function and inflammation. *Prog Neuropsychopharmacol Biol Psychiatry* 2022;**118**:110576.
33. Cacheiro-Llaguno C, Mösges R, Calzada D, González-de la Fuente S, Quintero E, Carnés J. Polysensitisation is associated with more severe symptoms: the reality of patients with allergy. *Clin Exp Allergy* 2024;**54**:607–20.