

MR MICHAEL SALVERMOSER (Orcid ID : 0000-0001-9412-8483)

DR ANDREAS BÖCK (Orcid ID : 0000-0002-4511-7769)

Article type : Original Article-Asthma and Rhinitis

Childhood asthma: novel endotyping by cytokines, validated through sensitization profiles and clinical characteristics

Salvermoser M^{1*}, Zeber K^{1*}, Boeck A¹, Klucker E², Schaub B^{1,3,#}

¹ Department of Pulmonary and Allergy, Dr. von Hauner Children's Hospital, LMU University of Munich, Germany

² Dr. von Hauner Children's Hospital, LMU University of Munich, Germany

³ Member of CPC, German centre of lung research, Munich, Germany

* Shared first authors

Corresponding author

Correspondence

Bianca Schaub, Department of Pulmonary and Allergy, Dr. von Hauner Children's Hospital, LMU, Munich, Germany. Email: bianca.schaub@med.uni-muenchen.de

Funding information

BMBF 01GL1742A (AB, KaZ); DFG SCHA-997/7-1 (AB, BS, KaZ); DFG SCHA-997/8-1 (BS, MS); DFG SFB-TR22 (BS, EK); Else Kröner-Fresenius-Stiftung (AB, BS); EUROIMMUN AG - a PerkinElmer company, Lübeck, Germany (assays); Hauner Verein (KaZ, BS).

Contributors

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/CEA.13858](#)

This article is protected by copyright. All rights reserved

MS and KaZ analysed and interpreted the data and wrote the manuscript. AB contributed to data analysis. EK and KaZ recruited patients. BS was responsible for study design, data analysis, interpretation and final manuscript version.

Data availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Condensed title

Childhood asthma: novel endotyping by cytokines

Key words

Cytokine; Endotypes; Validation; Unsupervised learning; Childhood allergy; Childhood asthma; Sensitization

Abstract

Background: Specific allergy sensitization pattern, using “component-resolved diagnosis” (CRD), is a central component of allergy and asthma in childhood. Besides this, allergic asthma has been characterized by a Th2-shifted endotype with elevation of classical Th2 cytokines. Recently, other endotypes with distinct mechanisms focusing on cytokine regulation evolved, yet those pathways are still not well understood.

Objective: (1) To define reproducible immunological endotypes using cytokine expression in an asthma cohort and (2) to characterize their sensitization profile and clinical phenotype.

Methods: Supernatants from PBMCs of 234 children (median age 10 years) of an asthma cohort were analysed for cytokine expressions. The children were split into a training (n=49) and validation (n=185) group. The training group was used to identify immunological endotypes by clustering cytokine expressions, which were then assessed regarding clinical characteristics and specific IgE of recombinant allergen components. Next, our findings were validated in the validation group.

Results: We identified novel endotypes based on primarily unstimulated cytokine expression. One endotype showed an IFN- γ /IL-17/IL-5 predominance, a different sensitization pattern (high in birch/apple; $p < 0.01$), and inferior lung function ($p < 0.01$). A second endotype grouped young children with food allergy and reduced lung function. Our findings were reproducible in the validation group.

Conclusion and clinical relevance: We identified two novel clinical asthma endotypes via cytokine expression pattern with distinct sensitization patterns. These novel findings are critical for clinical guidance and open avenues for identifying underlying mechanisms and more patient-specific therapies.

Abbreviations

AIT	Allergen-specific immunotherapy
AR	Allergic rhinoconjunctivitis
CLARA/CLAUS	Clinical Asthma Research Association
CRD	Component-resolved diagnosis
FEV1	Forced Expiratory Pressure in 1 Second
FVC	Forced Vital Capacity
HDM	House dust mite
IFN	Interferon
IL	Interleukin
LpA	Lipid A
PBMC	Peripheral blood mononuclear cells
Ppg	Peptidoglycan
PR10-proteins	Pathogenesis related protein family 10
sIgE	specific Immunoglobulin E
Th	T-helper
TLR	Toll-like receptor

Introduction

Despite rising prevalence of allergies, novel approaches for patient-specific endotyping and therapy are still developing slowly but are important to provide optimal therapy to distinct patient groups. Specific allergy sensitization pattern and cytokine regulation are two central components of allergy development in childhood. Due to the recent rise of “component-resolved diagnosis” (CRD), allergy sensitization patterns can be described in greater detail as compared to sensitization tests based on natural extracts of allergen sources. CRD is based on major and minor allergen recombinants. Major allergens, which are usually responsible for clinical symptoms, are main component of allergen-specific immunotherapy (AIT). AIT is the only disease-modifying treatment for allergic conditions, which cannot only improve symptoms but also has a preventive effect on the development of asthma.[1,2]

There has been a lot of research in the CRD area with focus on distinguishing true sensitization versus cross reactivity, predictors for a successful immunotherapy in patients with allergic rhinoconjunctivitis (AR) and asthma as well as prevalence of sensitization patterns and changes during growing up, identifying a phenomenon termed “molecular spreading”. [3-8] Beside allergic sensitization, allergic asthma has been characterized by a Th2-shifted endotype with elevation of classical Th2 cytokines (IL-4, IL-5, IL-13). [9-11] New asthma therapies target IgE or Th2-cell-associated cytokines for example biologicals as Omalizumab, Mepolizumab or Dupilumab (all approved for children). [12] However, there are far more endotypes with different clinical presentations (phenotypes). [11,13] The distinct mechanistic pathways are not quite understood yet, cytokines, as immunomodulatory proteins, play an important role. [12,14-16] Cluster analyses using cytokines in sputum to identify possible asthma or chronic rhinosinusitis endotypes in adults were reported. [17-19] In children, cluster analysis based on plasma cytokines was used to explore complex relationships among cytokines and allergic disease such as asthma and atopic dermatitis. [20]

Asthma endotypes based on cytokine clustering in children have not been reported so far. However, more comprehensive analyses of distinct regulation of cytokines may unravel novel endotypes of childhood asthma and allergies.

In this study, we aimed to answer two main questions: Can we identify reproducible immunoregulatory endotypes based on cytokine expression in a childhood asthma cohort? Do those endotypes differ in their sensitization profile and can we assign them to clinical phenotypes? This study was based on the hypothesis that not a single pathway defines

endotypes of a complex disease like asthma. Moreover, we believe that the balance of pathways and their counter regulation are more valuable for identifying reproducible endotypes.

Therefore, we combined machine learning strategies: unsupervised learning with feature selection and cluster validation on a distinct set of children. This novel approach distinguishes our study from most other studies using machine learning techniques for endotyping.

Methods

Study population

From our CLARA/CLAUS cross-sectional asthma study, we selected 234 children (median age 10.0 years) with or without asthma based on availability of cytokine expression data.[21,22] These 234 children were split into a training (n=49) and a validation (n=185) group. The training group comprises all children with available recombinant IgE measurements (Fig.1). Detailed population characteristics with focus on asthma, AR, and related symptoms in the past 12 months were available from questionnaires and clinical visits (Tab.1). Briefly, the CLARA/CLAUS cohort comprises steroid-naïve, newly-diagnosed mild-to-moderate asthmatics and healthy controls (without any clinical allergic symptoms).[21,22] Informed oral and written consent was obtained from the parents for participation and blood collection. Approval was obtained from the local ethics board, LMU Munich, Germany (Nr. 379-08; 4th of August 2008). This study does conform to the Declaration of Helsinki.

Lung function

Pulmonary function tests were performed using Master Screen PFT and MasterScreen Body (Jaeger, Germany) equipped with the software Lab Manager V5 32.0 (CareFusion, Germany). From each body plethysmography at least 3 measurements were taken to assure reproducibility. All tests were performed prior bronchodilator application.

Serum specific IgE assays

SlgE (major recombinant and natural extract) for the following allergen sources was measured by EUROLINE (Euroimmun AG, Lübeck) according to the manufacturer instructions: birch, grass (containing timothy), house dust mites (HDM; D.pteronyssinus, D.farinae) and furry animals (dog, cat). SlgE against these major allergen components were measured: Betv1, Phlp1, Phlp5Derp1, Derp2, Derf1, Derf2, Canf1, Feld1, and Mald1. We defined positive sensitization using a more specific cut off: slgE $\geq$ 0.7 kU/l.

Cytokine measurement

PBMCs were isolated within 24 hours after blood draw, cultured in X-Vivo (48 hours; Lonza, Basel, Switzerland) unstimulated or stimulated with plate-bound anti-CD3 (3 μ g/ml) plus soluble anti-CD28 (1 μ g/ml), lipid A (LpA; 0.1 μ g/ml), or peptidoglycan (Ppg; 1mg/ml) at 37°C. Cytokine levels were determined in supernatants of PBMCs with the Human Cytokine-Multiplex-Assay-Kit (Bio-Rad) by using Luminex (Austin, Texas) as previously described.[21] Cytokine expressions of 49 children and 32 cytokine features were available. Cytokine features relates to a child's 32 measured cytokine expressions. Those were composed by 8 cytokines (Th1-regulation: IFN- γ ;

Th2- and Treg-regulation: IL-4, IL-5, IL-13, IL-10; Th17-regulation: IL-17; pro-inflammatory markers: IL-10, IL-13, IL-17, IL-1 β , TNF- α) and 4 stimulation conditions (unstimulated, stimulated with anti-CD3/CD28, LpA, Ppg) (8x4=32 features). Cytokine selection was based on prior literature and their potential relevance in asthma pathogenesis.[21] We used LpA and Ppg in order to include two stimuli of the innate immune system targeting TLR4- or TLR2-stimulation, respectively.

Cytokine clustering

We applied unsupervised learning (*k*-means clustering) to detect underlying immunoregulatory endotypes using cytokine expression data only. Data preprocessing consisted of data imputation, data transformation, and feature selection.[23,24] During data pre-processing 5 (out of 32) cytokine features were excluded from the cluster analysis, based on a high proportion (>10%) of missing values due to non-detects: IL-13 (unstimulated, LpA-stimulated, Ppg-stimulated) and TNF α (LpA-/Ppg-stimulated). The number of clusters was chosen by evaluating four common measures: the elbow-method, the average silhouette method, the gap-statistic, and the bootstrapped prediction strength.[25-27] It was fixed after visually assessing cluster partitions. We provided a detailed cluster description showing the importance of cytokine features for clustering (p-value; Mann-Whitney test). More details were provided in the supplement.

Cluster validation

Our validation approach was two-staged. First, we looked at the same children (training group) but used different measurements (clinical data and specific IgE values). For each allergen, the sensitization prevalence of clusters were compared using a chi-squared test. Across all allergens, clusters were compared using a permutation test based on distances of the children's sensitizations pattern.

Second, the clustering was validated by looking at different children (validation group) but the same measurement (cytokine expressions). I.e. we applied the established workflow to the validation group. Next, the cytokine profiles of the new fitted clusters in the validation group were compared to the cytokine expression profiles of the clusters from the training group. The ability to generalize was measured using prediction strength.[27]

Analysis software

Data analyses were performed using R, graphical outputs were generated using the *ggplot2*-package.[28,29]

Results

Description of study population

Prior to answering our main objective, we described our study population stratified in training and validation group (tab.1). Our training group was smaller than the validation group (n=49 vs n=185) and had higher proportions of allergic diseases like asthma, RK, and food allergy. This was reflected in differences between eosinophils and total IgE. The reason for these differences is caused by the group assignment. The training group comprised only children with serum specific IgE measurements. Those were measured in children if serum was available and allergies known. Overall, our data comprises children with a median age of 10 years. Of note, we found high correlations between sIgE against major recombinant allergens and their corresponding natural extract for all allergen sources from table 1 (data not shown).

Immunoregulatory endotypes

As main aim, we clustered cytokine expressions, validated resulting endotypes, and examined their correlation with allergy sensitization patterns and clinical characteristics.

Prior to clustering cytokine expressions, we used sparse clustering to select cytokine features with a high clustering potential. Those were predominantly unstimulated or Ppg-stimulated cytokines (bar-height in Fig.S1A). Thus, these two stimuli were linked to more distinct cluster partitions. We confirmed this observation by clustering within each stimulus (Fig.S1B). Here we found less overlap for unstimulated and Ppg-stimulated cytokines as compared to LpA- or CD3/CD28-stimulated cytokines.

Using the selected features and comparing the cluster partitions for two and three clusters, we decided to further analyse the three-cluster solution as the other was obtained by combining two clusters (Fig.2A). Hence, we detected three clusters with 10 (red), 19 (blue), and 20 (green) children, respectively. The red cluster was clearly separated from the other two, which overlapped.

When comparing the cluster's cytokine expressions, we identified the following: First, there were no significant differences in anti-CD3/CD28-stimulated cytokine expressions (Fig.S2). Of note, level of cytokine secretion was roughly 10-fold higher following anti-CD3/28-stimulation for all cytokines apart from IL-4. Second, the red cluster was always highest across all cytokine-stimulus combinations. We found the strongest differences (asterisks in Fig.3) when comparing red to blue (unstimulated: IFN- γ , IL-17, TNF- α ; Ppg-stimulated: IFN- γ , IL-17, IL-4; LpA-stimulated IFN- γ , IL-17) and to green (unstimulated: IL-5 and Ppg-stimulated: IL-5). Third, blue and green cluster were significantly different for IFN- γ (unstimulated, Ppg-stimulated, LpA-stimulated) and IL-5

(unstimulated). Fig.3 was obtained by reducing Fig.S2 to stimuli and features with strong differences. Of note, we found no differences in cytokine expressions when comparing samples taken during or out of allergic seasons (data not shown).

Validation I: Immunoregulatory endotypes showed differences in sensitization profiles and clinical characteristics

On the level of sensitizations against major allergen recombinants, we identified differences across the clusters (Fig.4). The red cluster (red) had a higher sensitization rate against Mald1 and Betv1 and lowest sensitization for all other allergen components. Blue was highest for HDM, and timothy, whereas green showed high sensitization against animals and lower sensitization for Derf2/p2, and Phlp5. On allergen level, we observed significant differences for Derf2 and Feld1 (Fig.4). Overall, the red cluster had a significant dissimilar sensitization pattern compared to the other clusters (permutation test on distances, $p=0.01$). Results were similar using a less sensitive sensitization cut-off of 0.35 kU/l (Fig.S3).

For patient's characteristics, we presented differences across endotypes (Tab.2). The share of patients with AR was lowest in the red cluster (30%, red vs blue $p=0.01$). The blue cluster had most AR (84%), followed by green (65%). The blue cluster had most perennial symptoms of an AR (32%). The green cluster showed a higher level of patients with food allergy (55%) as compared to red (40%) and blue (21%), yet not significant. The green cluster comprises younger children aged 7.2 years with lowest height and weight (all $p<0.05$).

The red cluster had a lower lung function compared to other clusters. However, this is only significant for FEV1 when compared to blue. The number of leukocytes (8.9 G/l) was highest in the green cluster ($p<0.01$). IgE levels were highest for green (446.5 IU/ml), followed by red (378.5 IU/ml) and blue (282.0 IU/ml). The prevalence of polysensitization (≥ 3) was highest in the blue (74%) and different to the green cluster ($p<0.01$). Clusters did not differ in amounts of eosinophil and neutrophil cells.

Validation II: Same immunoregulatory endotypes in the validation group

When validating the clustering results on 185 distinct children, the red cluster showed a high prediction strength (91.3%). I.e. nearly all children of the new fitted red cluster were predicted to be in the same cluster. The others had a lower prediction strength (green: 49.3%, blue: 51.4%) (Fig.2B). The new fitted clusters had a very similar shape compared to the predicted clusters and again the red cluster was more distinct whereas blue and green cluster had a greater overlap (grey polygons in Fig.2B).

Additionally, the new fitted red cluster showed a very similar cytokine expression pattern (Fig.5) as compared to the initial clustering on the training group (Fig.3). I.e. the new fitted red cluster had a high cytokine expression across cytokines and stimulation protocols. Mostly, the new fitted blue and green cluster showed the same trends as in the training group (Fig.3). These trends tended to get more precise as the sample size increased. All cytokine features of the validation group were shown in the supplement (Fig.S4).

Discussion

In this study, we identified and validated a novel endotyping based on cytokine patterns in a steroid-naïve newly-diagnosed paediatric asthma cohort. These endotypes had different sensitization profiles and clinical characteristics. One endotype was most prominent with a reduced lung function and increased cytokine activation. Thus, a closer clinical supervision to this group in particular is necessary and may help to prevent exacerbations. These new findings open avenues for identifying critical underlying mechanisms for distinct groups of children and therefore help for more patient-specific therapies (e.g. anti-cytokine treatment).

Identification of novel endotypes

In this study, we presented a novel endotyping based on cytokine profiles of 49 children in an asthma cohort.[21,22]

Several groups reported cluster analyses using cytokines measured in sputum to identify possible asthma endotypes in adults.[17,18] Similar to our results, a previous study showed that classical Th2 cytokines (IL-4, IL-13) did not cluster, which underlines the inflammatory heterogeneity in asthma among different asthma patients.[17] Besides asthma, cytokine clustering was used to identify inflammatory endotypes of adult patients with chronic rhinosinusitis.[19]

Distinct to the adult studies, we identified new clusters based on cytokine-profiles in children with newly-diagnosed asthma, which is specific in our findings. Understanding cytokine profiling in children gives an insight into mechanism of disease development and possible options for therapeutic interventions in an early stage of disease. On a statistical level, reduction of random noise by feature selection and use of a validation group are distinguishing features of this study.[17-19]

Of note, our major findings were identified in unstimulated and Ppg-stimulated cells. Ppg is predominantly recognized by TLR2 and LpA by TLR4, which both activate the innate immune system.[30] Anti-CD3/CD28 stimulation activates the adaptive immune system by strongly stimulating T-cells[31], resulting in high T-cell-related cytokine secretion as observed in our study. This potent stimulation may override potential differences as identified in unstimulated cells and can explain missing differences in this condition. Confirmed in the larger group of children, this is critical for future application of clustering as this potent T-cell activation may be too strong for identifying differences in immune regulation. Furthermore, we examined both innate and adaptive-related cytokines. Differences in unstimulated and Ppg-stimulated cells were primarily associated with TNF- α as pro-inflammatory cytokine and Th1/Th2- and Th17-focused cytokines such as INF- γ , IL-17, and IL-5.

Heterogeneity in cytokine expression elucidate that a combination of differential T-cell regulation and not a single T-cell subset seems important for identifying these novel clusters and might represent various inflammatory subtypes of asthma. This also goes in line with our previous studies, which supported the critical role of regulatory T-cells in the immune-regulatory mechanisms of childhood asthma and using immunological data, genetic and environmental factors to classify childhood asthma endotypes.[21,32]

Clinical interpretation of novel endotypes

Analyzing the identified cytokine clusters in association with recombinant allergens, these clusters allowed to dissect three groups of children. Children in the red cluster, characterized by a higher apple (Mald1) and birch (Betv1) sensitization rate, were those with highest cytokine expression (INF- γ , IL-17, IL-5, TNF- α) and lowest leucocytes without any differences in eosinophils or neutrophils. A ubiquitous activation of several T-cell subsets, indicating an active diseases process, may be relevant for clinical practice. In line with this, an almost doubled level of IL-4 and TGF- β in patients sensitized to Betv1 compared with patients sensitized to D.pteronysinus was found.[33] Including clinical characteristics, the red cluster showed the lowest rhinoconjunctivitis rate. Lung function (z-scores of FEV1and FEV1/FVC) values were significantly lower as compared to the other two groups. In line with our data, asthmatics with high IL-5, IL-17A and IL-25 had significantly worse lung function parameters.[34] This group may be of special clinical relevance in order to control inflammation in parallel to reduced lung function previously not reported specifically. Yet, our data do not point to one target of treatment. Specifically, IL-17 is part of regulation, but not a central target. Confirming this, several studies in adult asthma have actually failed to be effective using anti-IL-17 antibodies.[35] However, we believe that a more in depth understanding of a child's immunoregulatory endotype may help selecting a more patient-specific therapy.

The second identified group (blue) included children with lowest level for IFN- γ , a Th1/Th17 cytokine, and the highest rates of AR and perennial symptoms. In parallel, they showed higher sensitization against HDM (Derf2, $p < 0.05$) and cat ($p < 0.05$). This mimics perennial sensitization possibly associated with reduction of Th1/Th17-regulation, potentially favoring a pro-allergic endotype, as previously suggested.[36] Consistent with our findings, a defective IFN- γ production by PBMC in a small group of HDM-allergic adults was detected, when stimulated with allergen (D.pteronysinus) and with an ubiquitous antigen.[37] A causal relation cannot be deduced by our data. Asthmatic children with perennial sensitization may have a reduced capacity of Th1 not able to counter regulate AR with lung function at normal levels. Yet, we propose it is important to study this group of children in a longitudinal design in order to identify this endotype prior to

symptomatic pattern and possibly evaluate options to treat them specifically with AIT during early phases of symptoms.

In the green cluster, children with intermediate cytokine patterns were clinically characterized by young age, more food allergy, and AR. This is in concordance with higher rates of food allergy in young children, also related to higher IgE levels. Of note, the immune pattern does not suggest an overall maturational difference between groups. Similar to the cytokine pattern, the sensitization pattern is intermediate except for highest sensitization to animals. Finally, these children seem to represent young asthmatics with food allergy predominance and reduced FEV1, also indicating the need for a closer clinical supervision for asthma control.

Validation of endotypes

We acknowledge the low number of patients in the cluster analysis (n=49). However, as we were able to replicate our findings in a larger validation group (n=185), we are convinced that our findings are robust and not subject to random noise (Fig.2B). Having a smaller training than validation group is unusual in machine learning. Since we did not rely on supervised learning algorithms but used an unsupervised clustering approach, we are convinced that this group distribution strengthens our results. Moreover, we believe that the underlying immunological mechanism is strong enough to be detected even in a smaller training group.

We were able to replicate our findings in a validation group from the same cohort. Especially the red cluster showed a high reproducibility expressed in a high prediction strength (91.3%). As already indicated by the cluster merging and overlap and overlap (Fig.2A-B), the blue and green clusters are more difficult to disentangle. However, their representative cytokine patterns were reproducible as well (Fig.5).

Our validation group had a different mixture of clinical phenotypes, however the number of asthmatic children exceeded the number of asthmatics in our training group (48 vs 76). The different phenotype mixture in our validation cohort in combination with replication of findings, we interpret as strength of this study. This might indicate that the immunological endotypes are not exclusive for the disease of asthma and might be important in other diseases as well. Our endotypes showed different sensitization patterns and clinical characteristics. Therefore, we believe that these immunological endotypes might be beneficial for a more patient specific clinical guidance.

Strengths and limitations

Of note, some facets of this study are important to be addressed for future approaches. While our analyses were based on limited number of cases in some subgroups, we could characterize them

in depth regarding sensitization and cytokine regulation. Yet, the study design did not allow nasal provocation test to evaluate specific allergies nor did we report treatment for AR in sufficient depth and related symptoms could not be causally related as this question was not included. Due to ethical reasons, it is not possible to perform bronchoalveolar lavages or lung biopsies for children regularly to measure cytokine expressions locally. Furthermore, there were no longitudinal data available to check consistency of these immunological endotypes over time. Yet, we believe this study is important as it identified novel clusters of immune regulation confirmed in a larger group of children in parallel with sensitization pattern and clinical asthma severity. For future treatment strategies, it is critical that unstimulated cytokine regulation not requiring detailed laboratory set-ups is able to differentiate these endotypes. If future additional studies can overcome the challenges of sample size and provocation tests, implementation into clinical practice may be truly relevant for asthma treatment.

Conclusion

This study represented a novel endotyping approach based on cytokine profiles and validated them on sensitization pattern, clinical characteristics, and in a group of distinct children. This study identified in particular two novel groups of children, namely those with a higher sensitization to Betv1 and Mald1 in parallel with systemic inflammation and reduced lung function and the group of young children with food allergy in parallel to reduced lung function both requiring a closer clinical follow-up and specific therapy as previously performed. This novel asthma endotyping thus offers new avenues not only for research purposes but most important for clinical guidance of this important paediatric population of manifest asthmatics.

Acknowledgements

We would like to thank all participating families. Special thank also to Tatjana Netz for her technical support and the study physicians for recruiting patients and collecting blood samples.

Conflict of Interest

Bianca Schaub reports grants from DFG, grants from EU, during the conduct of the study; grants from DFG, personal fees from Consultancy, outside the submitted work. Elisabeth Klucker reports personal fees from EUROIMMUN AG during the conduct of the study. Michael Salvermoser, Kathrin Zeber, and Andreas Boeck have nothing to disclose.

References

1. Halcken S. Allergen Immunotherapy for the Prevention of Allergic Asthma. Current Treatment Options in Allergy. 2018;5.
2. He YT, Zhou Y, Shao Q, Gan C, Chen M, Bao YL, et al. Immunoregulatory Effects of Subcutaneous Immunotherapy on Lymphocyte Subgroups and Cytokines in Children with Asthma. J Immunol Res. 2019;2019:7024905.
3. Lee JH, Kim SC, Choi H, Jung CG, Ban GY, Shin YS, et al. A Retrospective Study of Clinical Response Predictors in Subcutaneous Allergen Immunotherapy With House Dust Mites for Allergic Rhinitis. Allergy Asthma Immunol Res. 2018;10(1):18-24.
4. Douladiris N, Savvatanos S, Roumpedaki I, Skevaki C, Mitsias D, Papadopoulos NG. A molecular diagnostic algorithm to guide pollen immunotherapy in southern Europe: towards component-resolved management of allergic diseases. Int Arch Allergy Immunol. 2013;162(2):163-72.
5. Canonica GW, Ansotegui IJ, Pawankar R, Schmid-Grendelmeier P, van Hage M, Baena-Cagnani CE, et al. A WAO - ARIA - GA²LEN consensus document on molecular-based allergy diagnostics. The World Allergy Organization journal. 2013;6(1):17.
6. Dramburg S, Matricardi PM. Molecular Diagnosis of Allergy: The Pediatric Perspective. Front Pediatr. 2019;7:369.
7. Hatzler L, Panetta V, Lau S, Wagner P, Bergmann RL, Illi S, et al. Molecular spreading and predictive value of preclinical IgE response to Phleum pratense in children with hay fever. J Allergy Clin Immunol. 2012;130(4):894-901 e5.
8. Geroldinger-Simic M, Zelniker T, Aberer W, Ebner C, Egger C, Greiderer A, et al. Birch pollen-related food allergy: clinical aspects and the role of allergen-specific IgE and IgG4 antibodies. J Allergy Clin Immunol. 2011;127(3):616-22 e1.
9. Foster PS, Maltby S, Rosenberg HF, Tay HL, Hogan SP, Collison AM, et al. Modeling TH 2 responses and airway inflammation to understand fundamental mechanisms regulating the pathogenesis of asthma. Immunological reviews. 2017;278(1):20-40.
10. Schatz M, Rosenwasser L. The allergic asthma phenotype. The Journal of Allergy and Clinical Immunology: In Practice. 2014;2(6):645-8.
11. Kuruvilla ME, Lee FE, Lee GB. Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease. Clinical reviews in allergy & immunology. 2019;56(2):219-33.
12. Lambrecht BN, Hammad H, Fahy JV. The Cytokines of Asthma. Immunity. 2019;50(4):975-91.

13. Fitzpatrick AM, Bacharier LB, Jackson DJ, Szeffler SJ, Beigelman A, Cabana M, et al. Heterogeneity of Mild to Moderate Persistent Asthma in Children: Confirmation by Latent Class Analysis and Association with 1-Year Outcomes. *J Allergy Clin Immunol Pract*. 2020.
14. Pawankar R, Hayashi M, Yamanishi S, Igarashi T. The paradigm of cytokine networks in allergic airway inflammation. *Curr Opin Allergy Clin Immunol*. 2015;15(1):41-8.
15. Bhakta NR, Woodruff PG. Human asthma phenotypes: from the clinic, to cytokines, and back again. *Immunol Rev*. 2011;242(1):220-32.
16. Vizmanos-Lamotte G, Moreno-Galdó A, Muñoz X, Gómez-Ollés S, Gartner S, Cruz MJ. Induced sputum cell count and cytokine profile in atopic and non-atopic children with asthma. *Pediatr Pulmonol*. 2013;48(11):1062-9.
17. Seys SF, Scheers H, Van den Brande P, Marijsse G, Dilissen E, Van Den Bergh A, et al. Cluster analysis of sputum cytokine-high profiles reveals diversity in T(h)2-high asthma patients. *Respir Res*. 2017;18(1):39.
18. Hinks TSC, Brown T, Lau LCK, Rupani H, Barber C, Elliott S, et al. Multidimensional endotyping in patients with severe asthma reveals inflammatory heterogeneity in matrix metalloproteinases and chitinase 3-like protein 1. *J Allergy Clin Immunol*. 2016;138(1):61-75.
19. Tomassen P, Vandeplas G, Van Zele T, Cardell L-O, Arebro J, Olze H, et al. Inflammatory endotypes of chronic rhinosinusitis based on cluster analysis of biomarkers. *Journal of Allergy and Clinical Immunology*. 2016;137(5):1449-56. e4.
20. Yamamoto-Hanada K, Kawakami E, Saito-Abe M, Sato M, Mitsubuchi H, Oda M, et al. Exploratory analysis of plasma cytokine/chemokine levels in 6-year-old children from a birth cohort study. *Cytokine*. 2020;130:155051.
21. Raedler D, Ballenberger N, Klucker E, Bock A, Otto R, Prazeres da Costa O, et al. Identification of novel immune phenotypes for allergic and nonallergic childhood asthma. *J Allergy Clin Immunol*. 2015;135(1):81-91.
22. Krautenbacher N, Flach N, Böck A, Laubhahn K, Laimighofer M, Theis FJ, et al. A strategy for high-dimensional multivariable analysis classifies childhood asthma phenotypes from genetic, immunological, and environmental factors. 2019;74(7):1364-73.
23. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics (Oxford, England)*. 2012;28(1):112–8.
24. Witten DM, Tibshirani R. A framework for feature selection in clustering. *Journal of the American Statistical Association*. 2010;105(490):713–26.

25. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2001;63(2):411–23.
26. Kaufman L, Rousseeuw PJ. *Finding Groups in Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc; 1990.
27. Tibshirani R, Walther G. Cluster Validation by Prediction Strength. *Journal of Computational and Graphical Statistics*. 2005;14(3):511–28.
28. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria; 2019.
29. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag New York; 2009.
30. Schaub B, Bellou A, Gibbons FK, Velasco G, Campo M, He H, et al. TLR2 and TLR4 stimulation differentially induce cytokine secretion in human neonatal, adult, and murine mononuclear cells. *Journal of Interferon & Cytokine Research*. 2004;24(9):543–52.
31. Trickett A, Kwan YL. T cell stimulation and expansion using anti-CD3/CD28 beads. *Journal of Immunological Methods*. 2003;275(1–2):251–5.
32. Krautenbacher N, Flach N, Bock A, Laubhahn K, Laimighofer M, Theis FJ, et al. A strategy for high-dimensional multivariable analysis classifies childhood asthma phenotypes from genetic, immunological, and environmental factors. *Allergy*. 2019;74(7):1364–73.
33. Tyurin YA, Lissovskaya SA, Fassahov RS, Mustafin IG, Shamsutdinov AF, Shilova MA, et al. Cytokine profile of patients with allergic rhinitis caused by pollen, mite, and microbial allergen sensitization. *Journal of Immunology Research*. 2017;2017.
34. Seys S, Grabowski M, Adriaensen W, Decraene A, Dilissen E, Vanoirbeek J, et al. Sputum cytokine mapping reveals an ‘IL-5, IL-17A, IL-25-high’ pattern associated with poorly controlled asthma. *Clinical & Experimental Allergy*. 2013;43(9):1009–17.
35. Busse WW, Holgate S, Kerwin E, Chon Y, Feng J, Lin J, et al. Randomized, double-blind, placebo-controlled study of brodalumab, a human anti-IL-17 receptor monoclonal antibody, in moderate to severe asthma. *Am J Respir Crit Care Med*. 2013;188(11):1294–302.
36. Fenoglio D, Ferrera A, Ferrera F, Sormani M, Di Gioacchino M, Ciprandi G. Patients with allergic rhinitis show an allergen-specific interferon-gamma defect. *European Journal of Inflammation*. 2008;6(2):87–91.
37. McHugh S, Wilson A, Deighton J, Lachmann P, Ewan P. The profiles of interleukin (IL)-2, IL-6, and interferon-gamma production by peripheral blood mononuclear cells from house-dust-mite-allergic patients: a role for IL-6 in allergic disease. *Allergy*. 1994;49(9):751–9.

Accepted Article

Tab.1: Patients characteristics of training group used for clustering and validation group.

		Training group (n=49)		Validation group (n=185)		p-value
Asthma		48	(98.0%)	76	(41.1%)	<0.01
Rhinoconjunctivitis		32	(66.7%)	26	(14.4%)	<0.01
Food allergy		19	(38.8%)	22	(12.2%)	<0.01
Sex	female	20	(40.8%)	81	(43.8%)	0.83
Age	[years]	9.3	[6.6;11.5]	10.2	[6.4;12.1]	0.59
Height	[cm]	138	[119;151]	140	[120;154]	0.4
Weight	[kg]	31.5	[23.2;44.0]	33.5	[23.0;46.2]	0.42
BMI	[kg/m ²]	17.1	[15.7;18.3]	17.4	[15.5;19.7]	0.48
Breast feeding		44	(93.6%)	155	(85.6%)	0.22
	[months]	6.0	[3.0;8.8]	6.0	[2.0;10.0]	0.56
Siblings	younger	20	(40.8%)	81	(46.3%)	0.6
	older	19	(38.8%)	78	(44.6%)	0.58
Paternal	asthma	6	(12.5%)	21	(11.9%)	1.00
	eczema	1	(2.1%)	9	(5.1%)	0.69
	hay fever	16	(33.3%)	38	(21.6%)	0.13
Maternal	asthma	9	(18.4%)	16	(9.1%)	0.12
	eczema	5	(10.2%)	21	(12.0%)	0.92
	hay fever	13	(26.5%)	44	(25.3%)	1.00
Smoking		2	(4.2%)	31	(17.3%)	0.04
Pet at any time	dog	5	(10.2%)	26	(14.1%)	0.64
	cat	7	(14.3%)	36	(19.5%)	0.53
FEV1[†]	[%]	92	[84;101]	94	[85;102]	0.16
	[z-score]	-0.94	[-1.57;0.01]	-0.82	[-1.54;-0.18]	0.74
FVC[†]	[%]	98	[89;109]	98	[91;103]	0.42
	[z-score]	-0.06	[-0.82;0.98]	-0.18	[-0.70;0.28]	0.20
FEV1/FVC[†]	[%]	93	[83;106]	95	[90;100]	0.67
	[z-score]	-1.25	[-2.39;0.04]	-0.99	[-1.86;-0.47]	0.81
Leucocytes	[G/l]	8.1	[6.2;9.3]	7.3	[6.2;8.8]	0.22
Neutrophils	[% leucocytes]	46.9	[41.1;52.0]	47.0	[41.5;56.8]	0.33
	[G/l]	3.5	[2.9;4.5]	3.4	[2.6;4.6]	0.64
Eosinophils	[% leucocytes]	8.0	[5.0;11.0]	3.0	[2.0;7.0]	<0.01
	[G/l]	0.6	[0.4;0.8]	0.2	[0.1;0.5]	<0.01
IgE	[IU/ml]	338	[181;765]	86	[29;259]	<0.01
Apple Mal d1[‡]		7	(14.3)	-	-	-
Birch Bet v1[‡]		23	(46.9%)	-	-	-
Timothy Phl p1[‡]		20	(40.8%)	-	-	-
Timothy Phl p5[‡]		15	(30.6%)	-	-	-
D.pteronyssinus Der p1[‡]		16	(32.7%)	-	-	-
D.pteronyssinus Der p2[‡]		21	(42.9%)	-	-	-
D.farinae Der f1[‡]		21	(42.9%)	-	-	-
D.farinae Der f2[‡]		24	(49.0%)	-	-	-

Dog Can f1[‡]	3	(6.1%)	-	-	-
Cat Fel d1[‡]	27	(55.1%)	-	-	-

Categorical variables were shown as absolute values with percentages in brackets.

Numerical values were summarized by their median and interquartile range [Q0.25; Q0.75].

P-values originated from a Chi-squared test (categorical variable) or a Mann-Whitney test (continuous variables).

† lung function is shown solely for asthma patients

‡ specific IgE ≥ 0.7 kU/l, only available for the training group

Tab.2: Clinical and population characteristics of the training group stratified by cluster.

Training group (n=49)		Red (10)	Blue (19)	Green (20)	Differences [†]
Rhinoconjunctivitis		30%	84%	65%	Red vs Blue [§]
Asthma		90%	100%	100%	-
Food Allergy		40%	21%	55%	-
Time of symptoms	early	10%	5%	15%	-
	both	10%	32%	35%	
	late	0%	11%	5%	
	always	10%	32%	5%	
	never	70%	21%	20%	
	missing	0%	0%	20%	
Sex	female	50%	37%	40%	-
Age	[years]	10.5 [7.6;12.5]	11.4 [7.0;12.5]	7.2 [5.5;9.8]	Red/Blue vs Green [‡]
Height	[cm]	144 [127;161]	149 [124;156]	127 [114;140]	Red/Blue vs Green [‡]
Weight	[kg]	36.5 [27.1;44.6]	41.3 [24.6;45.4]	26.1 [20.4;31.6]	Blue vs Green [‡]
BMI	[kg/m ²]	17.1 [16.2;18.2]	17.3 [16.1;18.4]	16.1 [15.4;17.9]	-
Breast feeding		80%	100%	85%	-
	[months]	6 [5;7]	6 [4;9]	5 [1;9]	-
Siblings	younger	40%	37%	45%	-
	older	40%	32%	45%	-
Paternal	asthma	20%	5%	15%	-
	eczema	0%	5%	0%	-
	hay fever	40%	32%	30%	-
Maternal	asthma	30%	21%	10%	-
	eczema	10%	5%	15%	-
	hay fever	30%	32%	20%	-
Smoking		10%	5%	15%	-
Pet at any time	dog	10%	11%	10%	-
	cat	0%	21%	15%	-
FEV1	[%]	79 [78;93]	97 [89;106]	89 [82;102]	Red vs Blue [‡]
	[z-score]	-1.9 [-2.1;-1.3]	-0.3 [-1.0;0.3]	-1.2 [-1.7;-0.3]	Red/Green vs Blue [‡]
FVC	[%]	93 [87;112]	103 [96;111]	95 [88;105]	-
	[z-score]	-0.2 [-1.3;0.9]	0.5 [-0.4;1.1]	-0.3 [-0.9;0.6]	-
FEV1/FVC	[%]	86 [83;108]	93 [88;106]	97 [85;106]	-
	[z-score]	-2.0 [-2.6;0.7]	-1.1 [-2.2;0.1]	-1.1 [-2.2;-0.2]	-
Leukocytes	[G/l]	6.3 [5.9;8.2]	7.6 [6.2;8.5]	8.9 [8.2;9.7]	Red/Blue vs Green [‡]
Neutrophils	[% leucocytes]	46.0 [45.0;50.7]	48.0 [41.1;51.5]	46.5 [40.5;52.3]	-
	[G/l]	2.8 [2.7;3.9]	3.3 [2.9;4.0]	4.2 [3.3;4.7]	
Eosinophils	[% leucocytes]	8.0 [3.0 – 11.8]	8.0 [5.0 – 11.0]	7.0 [6.0 – 9.0]	-
	[G/l]	0.6 [0.2;0.8]	0.6 [0.4;0.8]	0.76 [0.4;0.8]	
IgE	[IU/ml]	379 [193;745]	282 [188;546]	447 [178;920]	-
Sensitizations natural extract	0	20%	5%	5%	Blue vs Green [§]
	1	40%	21%	25%	
	2	10%	0%	40%	
	≥3	30%	74%	30%	

Categorical variables were shown as percentages.

Numerical values were summarized by their median and interquartile range [Q0.25; Q0.75].

P-values originated from a Chi-squared test (categorical variable) or a Mann-Whitney test (continuous variables).

† p-value<0.05 ‡ Pearson's Chi-squared test § Mann-Whitney test

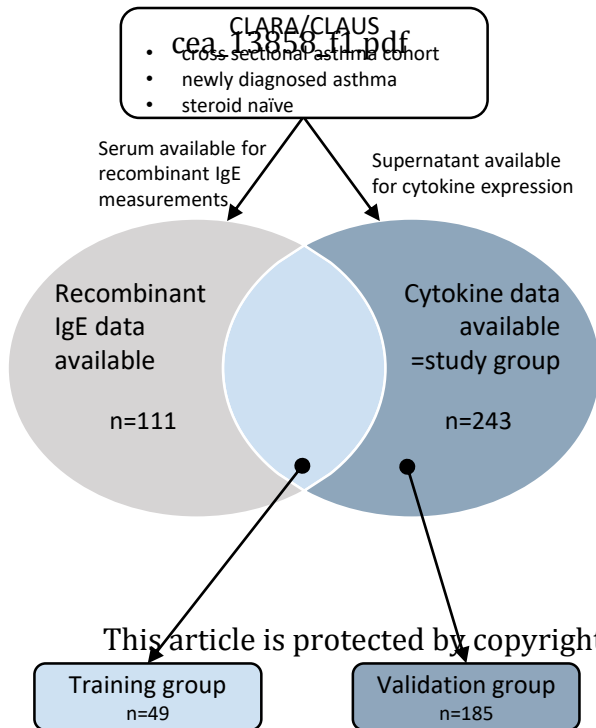
Fig.1: Flowchart of study population with train-validation split.

Fig.2: A Cluster partitions for both values of k on the first two principal components (explained variation as percentage). For $k = 3$ the clusters were colored with centroids (grey triangles) and the cluster-combining was indicated by a black polygon around blue and green clusters. **B** Cluster validation on validation group ($n=185$); point color and polygon indicating the **new fitted cluster**; background color representing the **predicted cluster**.

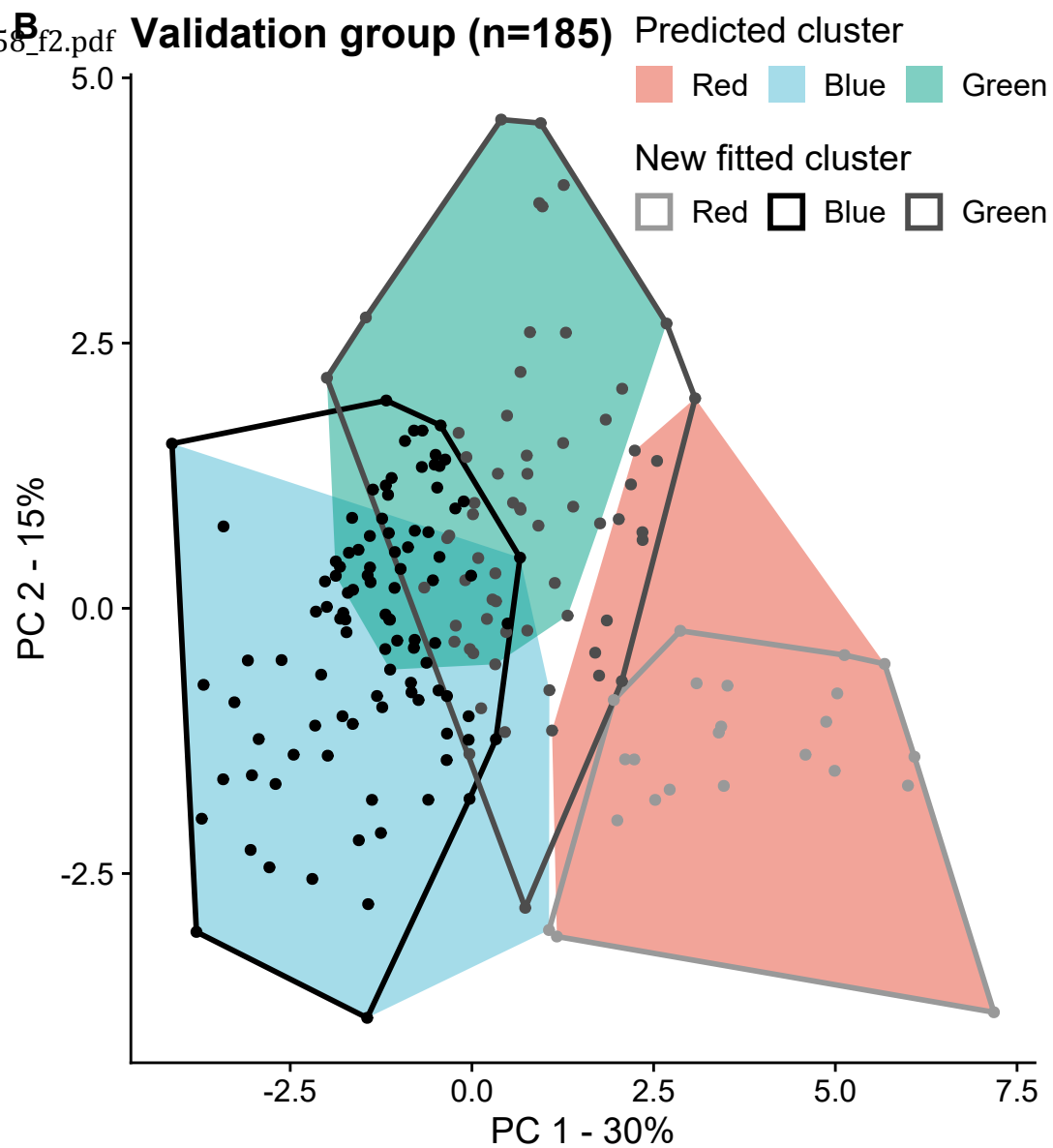
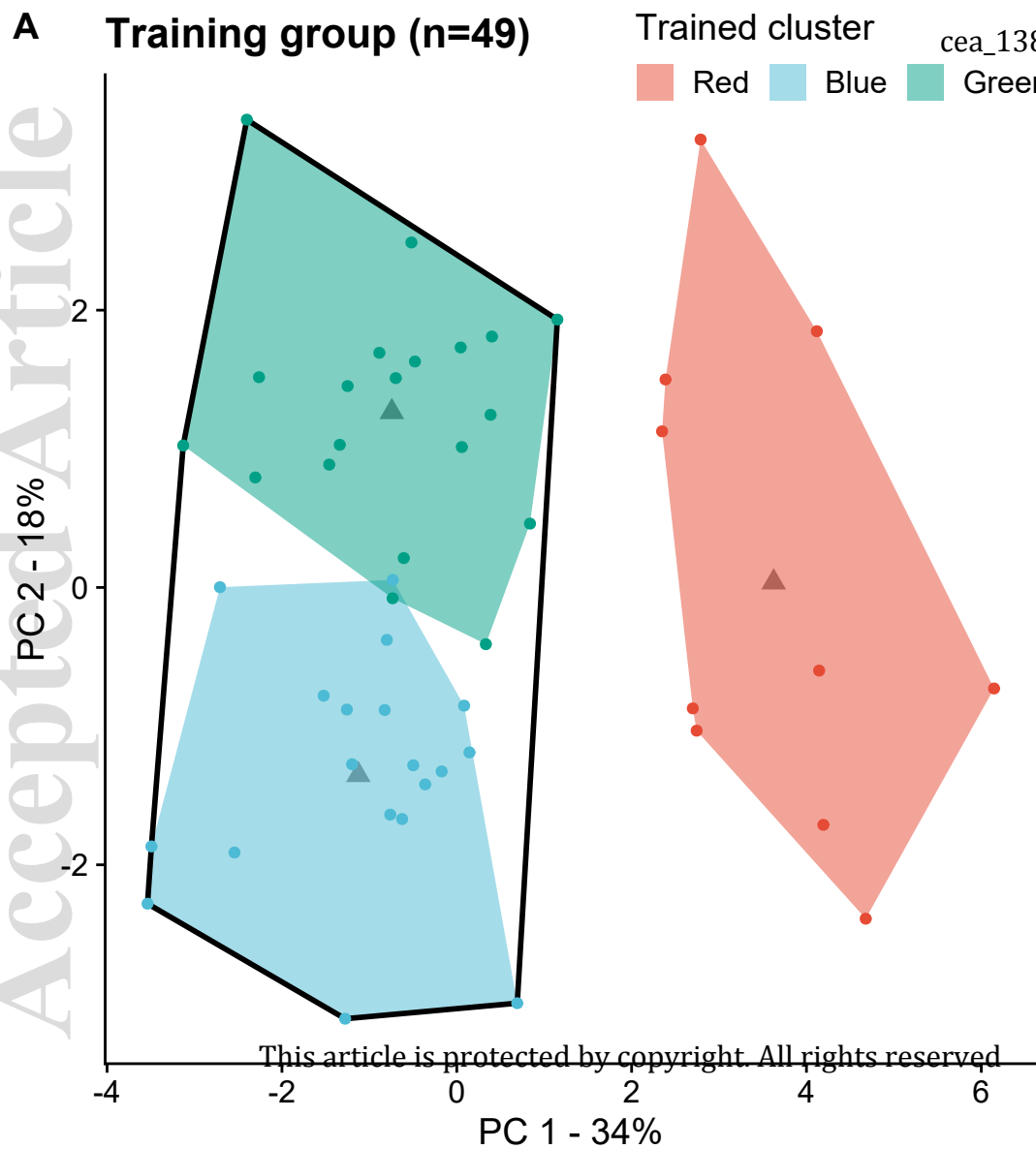
Fig.3: Study group ($n=49$): Differences across clusters (x-axis) for cytokine features (vertical panels) and stimulus (horizontal panels) on expression level (y-axis). The adjusted p-value measured a feature's effect on clustering; i.e. the lower the p-value the more impact on partitioning the endotypes.

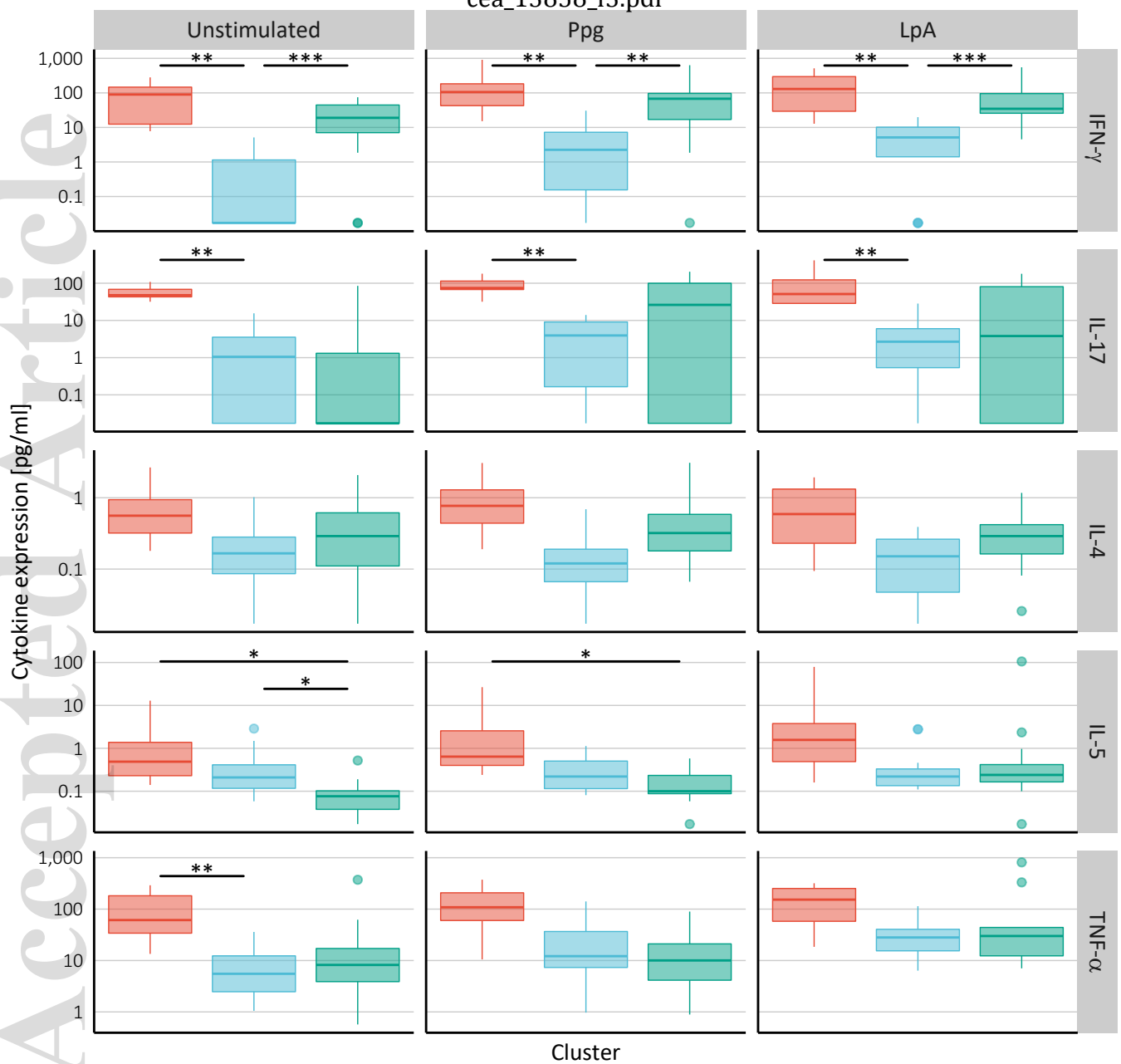
Fig.4: Sensitization rate (y-axis) for each cluster (x-axis) facetted by major allergen component. Significant pairwise Chi-squared tests on component level were indicated with horizontal bars. Overall, the sensitization pattern of the red cluster differed significantly from the others (permutation test: $p=0.01$). Sensitization was defined as specific IgE ≥ 0.7 kU/l.

Fig.5: Validation group ($n=185$): Differences across clusters (x-axis) for cytokine features (vertical panels) and stimulus (horizontal panels) on expression level (y-axis). The adjusted p-value measured a feature's effect on clustering; i.e. the lower the p-value the more impact on partitioning the endotypes.

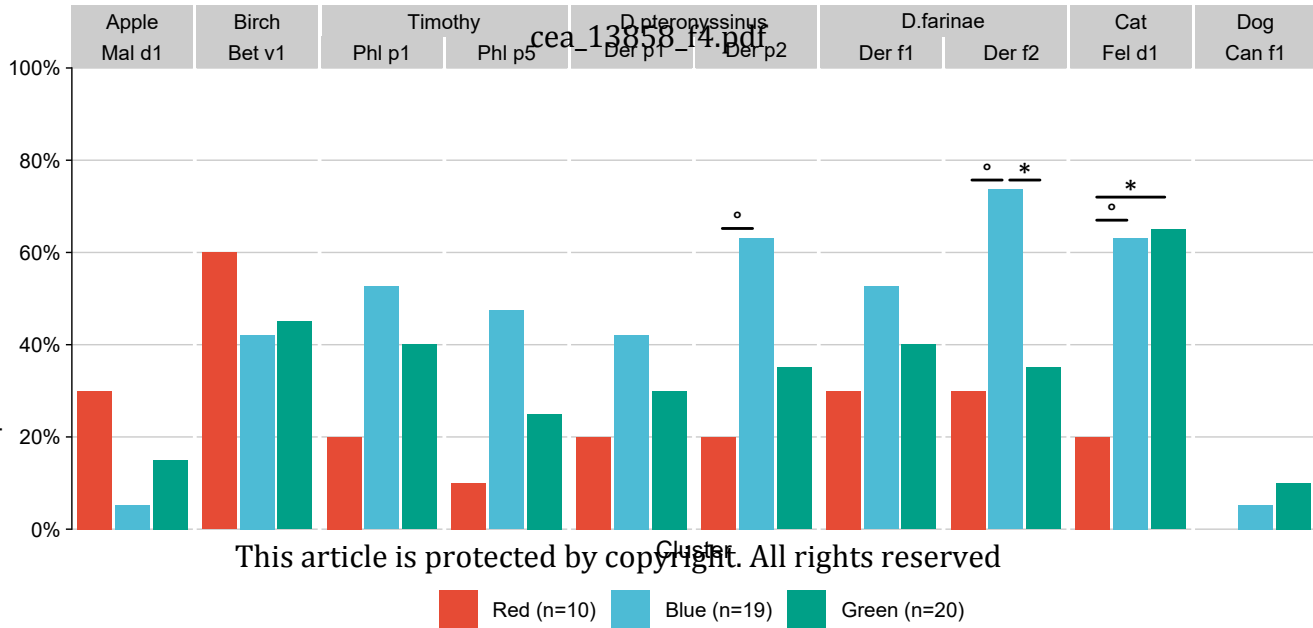


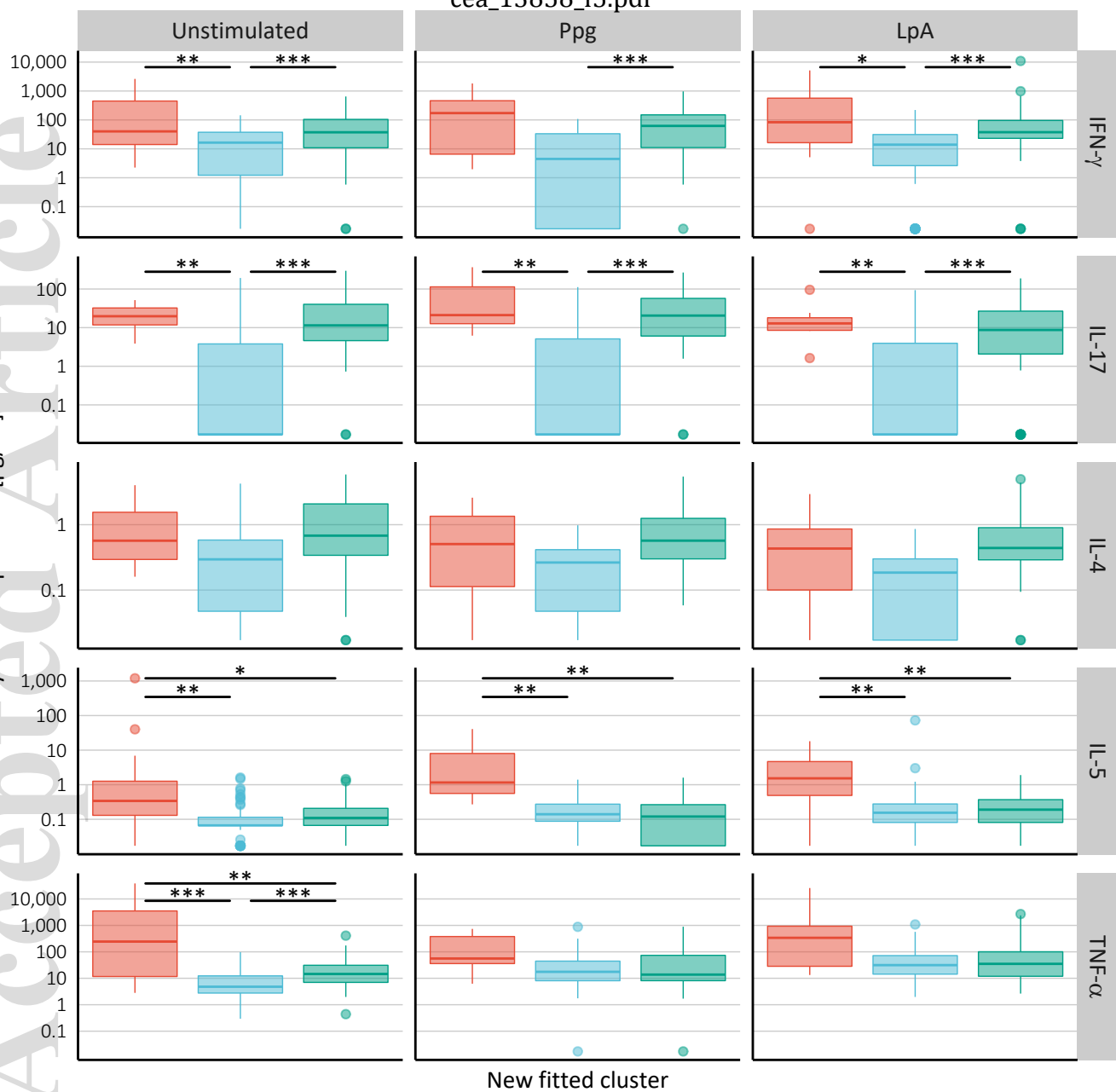
This article is protected by copyright. All rights reserved.





This article is protected by copyright. All rights reserved





Supplemental files

Childhood allergies: novel endotyping by cytokines, validated through sensitization profiles and clinical characteristics

Salvermoser M^{1*}, Zeber K^{1*}, Boeck A¹, Klucker E², Schaub B^{1, 3, #}

¹ Department of Pulmonary and Allergy, Dr. von Hauner Children's Hospital, LMU University of Munich, Germany

² Dr. von Hauner Children's Hospital, LMU University of Munich, Germany

³ Member of CPC, German centre of lung research, Munich, Germany

* Shared first authors

Corresponding author

Correspondence

Bianca Schaub, Department of Pulmonary and Allergy, Dr. von Hauner Children's Hospital, LMU, Munich, Germany. Email: bianca.schaub@med.uni-muenchen.de

Cluster analyses

Our aim was to detect underlying immunoregulatory patterns via cytokine expression data. Therefore, an unsupervised machine learning approach was necessary and k -means clustering suitable.

Before fitting the clustering algorithm, a few data preprocessing steps were necessary. First, we removed features with more than 10% missing values and imputed remaining missing values using a random forest approach.¹ Second, we normalized cytokine features by taking the 10th root, centered them around zero, and scaled them by their standard deviation. Then, we made use of Witten and Tibshirani's idea of a sparse clustering for feature selection.² They proposed a method that reduces random noise and carves out the underlying structure by removing features with little or no clustering strength. Their idea is based on weighting the features according to their ability to split the subjects in distinct clusters, whereas the weight of a feature reflects its clustering quality.

Then, we fitted a sparse k -means algorithm to the preprocessed data and selected features with a weight greater than the median weight (Fig.S2A).³ The filtering procedure put on average most weight on the unstimulated and Ppg-stimulated cytokine expression and least on the CD3/CD28-stimulated expression. Thus, unstimulated and Ppg-stimulated cytokine features were linked to a higher variation between clusters and improvement of clustering. This finding was consistent to fitting k -means to every stimulus itself (Fig.S2B). The features with the highest weights are unstimulated IL-1 β , unstimulated IL-17, and Ppg-stimulated IL-5. The median cut-off was chosen a priori as cutoff value in order to prevent data dredging by following a predetermined analysis pipeline.⁴

After preprocessing and feature selection, a major challenge in k -means clustering was determining an appropriate number of clusters k . As there is no best method for doing so, we used four popular approaches that compare the k -means algorithm for different values of k and synthesized the results. The within sum of squares function (elbow-method)⁵, the average silhouette method⁶, and the gap-statistic⁵ are focusing on the cluster quality as ratio of the intra cluster and between cluster variation. Additionally, a bootstrapped prediction strength measuring the algorithm's ability to predict cluster co-membership is evaluated.⁷ The results were shown in Fig.S5 A-D respectively. The elbow method (panel A) was monotonically decreasing and we saw some flattening around $k \in \{2,3\}$. The average silhouette (panel B) and gap method (panel C) provided a distinct optimal number of clusters by following an optimization rule, $k = 2$ and $k = 3$ respectively. The prediction strength (panel D) monotonously decreased, and preferred $k = 2$ over $k = 3$.

As there was no clear decision for using either two or three clusters, we fitted two k -means algorithm with 50 random starting points for both values. The resulting partitions were shown in Fig.3A. We found a clearly separated red cluster and two slightly overlapping clusters (blue and green). When reducing the number of clusters to $k = 2$, the red cluster stayed untouched and other clusters collapsed to a bigger cluster (gray convex hull in the background). As we found a clear separation of the red cluster for both values of k and some minor distinctions of the blue and green cluster, we chose $k = 3$. This entailed cluster sizes of 10 (red), 19 (blue), and 20 (green).

The clusters were described on cytokine expression level in Fig.S3. After pairwise testing the cytokine expressions and adjusting for multiple testing⁸, Fig.4 made use of the adjusted p-values as measure effect size for cluster separation and high effects (p-values < 0.05) were indicated with bars. I.e. a low p-value indicated high differences in clusters which triggers cluster distinction and vice versa. Fig.4 was the reduction of Fig.S3 to cytokines and stimuli with high effects across clusters.

Moreover, the clusters were described with their sensitization pattern (Fig.5) and basic clinical and epidemiological characteristics (Tab.1). Those served as sources of external validation, i.e. validating the cluster by variables not used for the clustering procedure. The sensitization patterns were compared on allergen level using chi-squared tests and on cluster level using a permutation test approach based on the average dissimilarity of sensitization patterns compared to the other clusters with 10000 random permutations.

Finally, we validated the clusters in terms of prediction strength on cytokine expression data of 188 distinct children from the same cohort (Fig.3B).⁷

75 **References**

- 76 1. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-
77 type data. *Bioinformatics (Oxford, England)*. 2012;28(1):112–118.
- 78 2. Witten DM, Tibshirani R. A framework for feature selection in clustering. *Journal of the*
79 *American Statistical Association*. 2010;105(490):713–726.
- 80 3. Witten DM, Tibshirani R. sparcl: Perform Sparse Hierarchical Clustering and Sparse K-Means
81 Clustering. In:2018.
- 82 4. Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nature*
83 *Human Behaviour*. 2017;1(1):e124.
- 84 5. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap
85 statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.
86 2001;63(2):411–423.
- 87 6. Kaufman L, Rousseeuw PJ. *Finding Groups in Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc;
88 1990.
- 89 7. Tibshirani R, Walther G. Cluster Validation by Prediction Strength. *Journal of Computational*
90 *and Graphical Statistics*. 2005;14(3):511–528.
- 91 8. Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under
92 Dependency. *The Annals of Statistics*. 2001;29(4):1165-1188.

93

94

95 **Fig.S1:** **A** Weights of sparse k-means clustering with median cut-off (red line) faceted by
96 stimulus. Features with weight above the cut-off were filled out. **B** Clustering of cytokine
97 features in each stimulus plotted on the first and second principal component (similar to
98 Fig.3A).

99 **Fig.S2:** Training group: Differences across clusters (x-axis) for cytokine features (horizontal
100 panels) and stimulus (vertical panels) on expression level (y-axis). Biggest differences were
101 indicated by a black bar.

102 **Fig.S3:** Sensitization rate (y-axis) for each cluster (x-axis) faceted by major allergen
103 component. Significant pairwise Chi-squared tests on component level were indicated with
104 horizontal bars. Overall, the sensitization pattern of the red cluster differed significantly from
105 the others (permutation test: $p=0.02$). Sensitization cut-off: 0.35 KU/l.

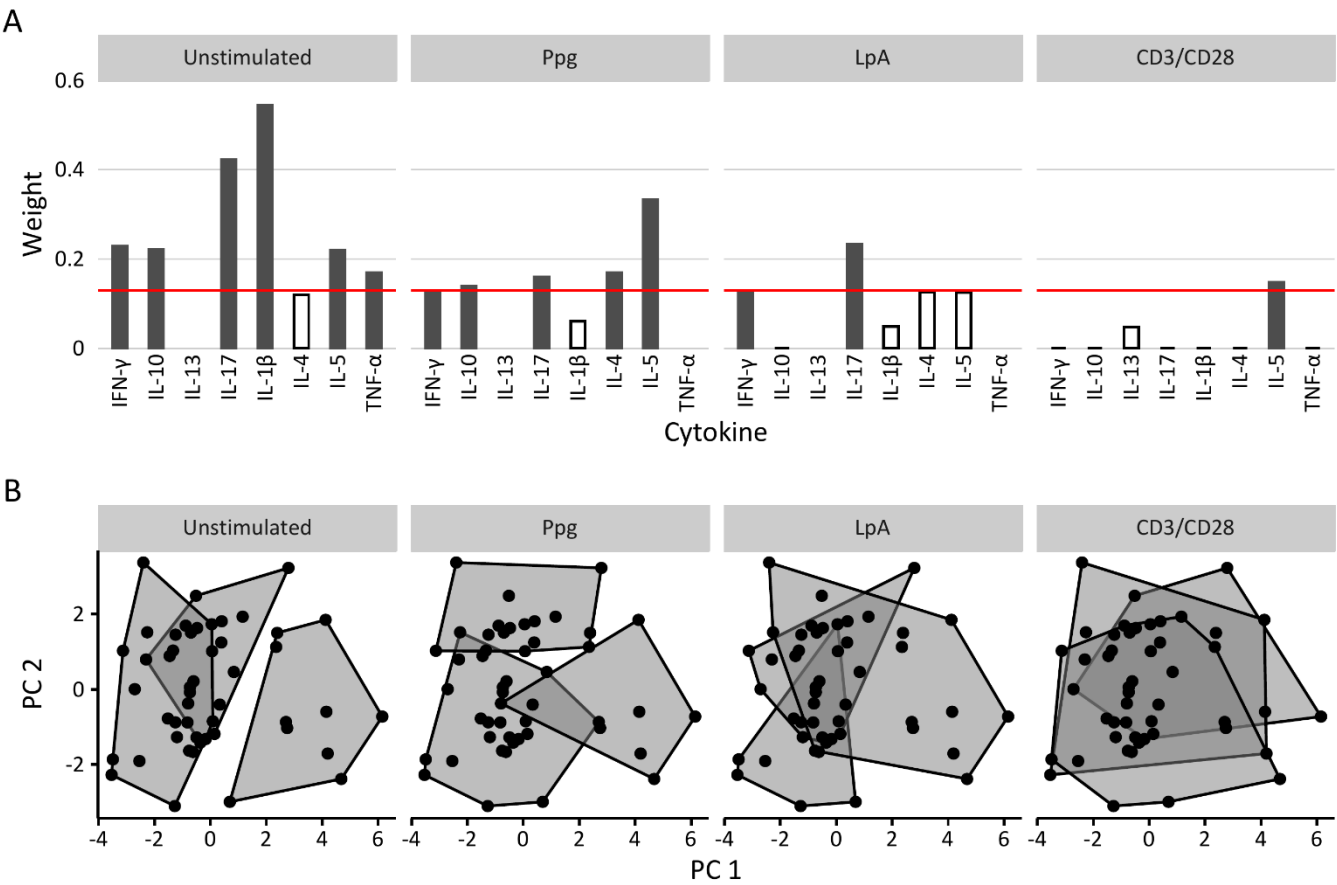
106 **Fig.S4:** Validation group: Differences across clusters (x-axis) for cytokine features (horizontal
107 panels) and stimulus (vertical panels) on expression level (y-axis). Biggest differences were
108 indicated by a black bar.

109 **Fig.S5:** Selection criteria (y-axis) for the number of clusters k (x-axis). The optimal choice was
110 indicated by blue background. It was not possible to provide an optimum for facet **A** as this
111 methods is based on the subjective identification of a knee ($k \in \{2, 3\}$).

112

113

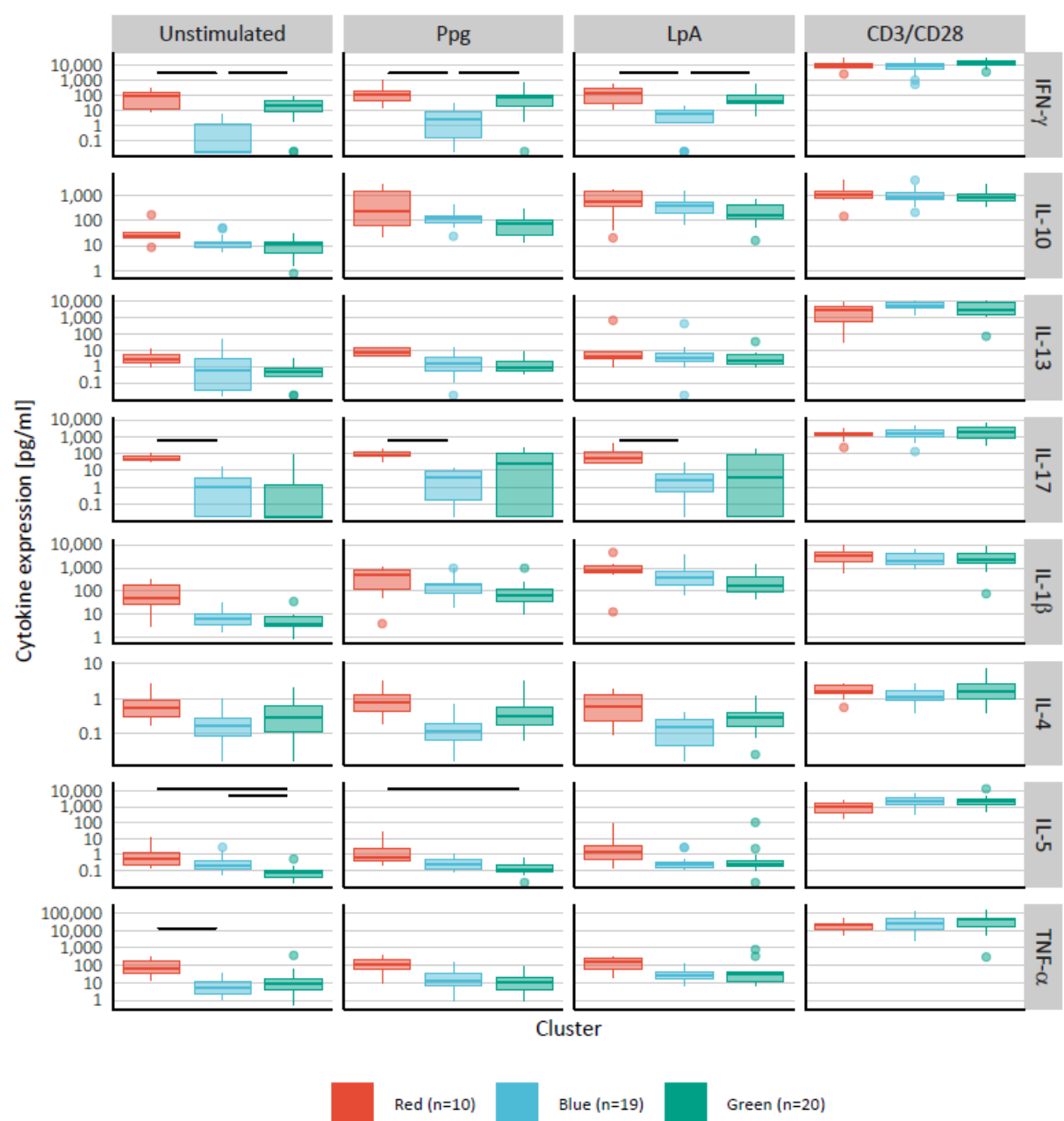
114



116

117

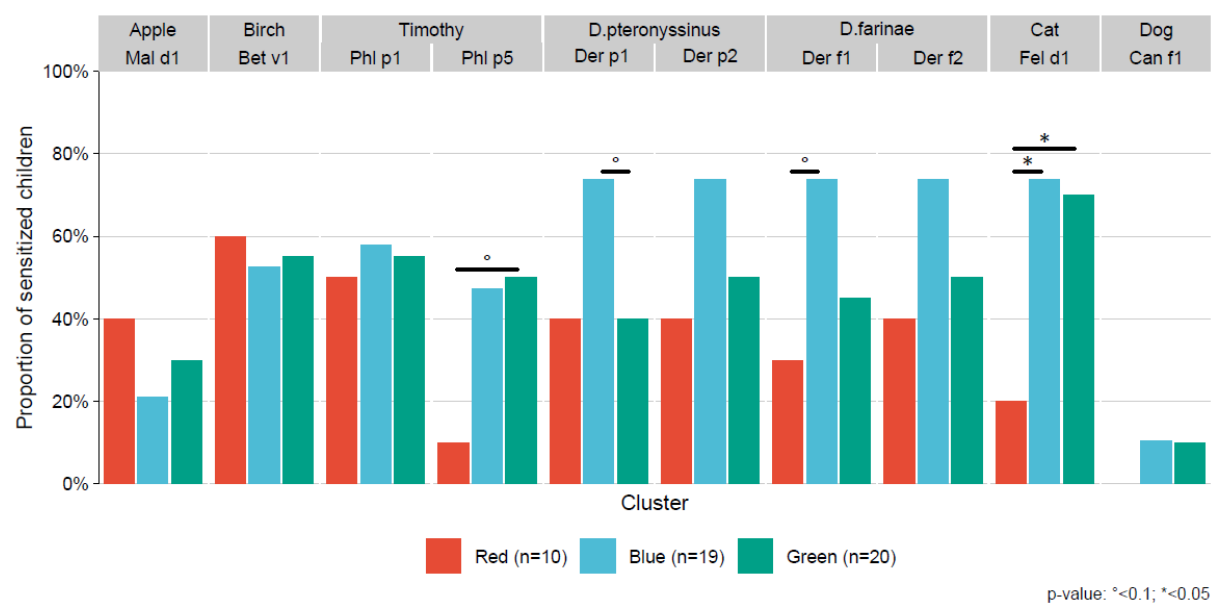
118 **Fig.S2**



119

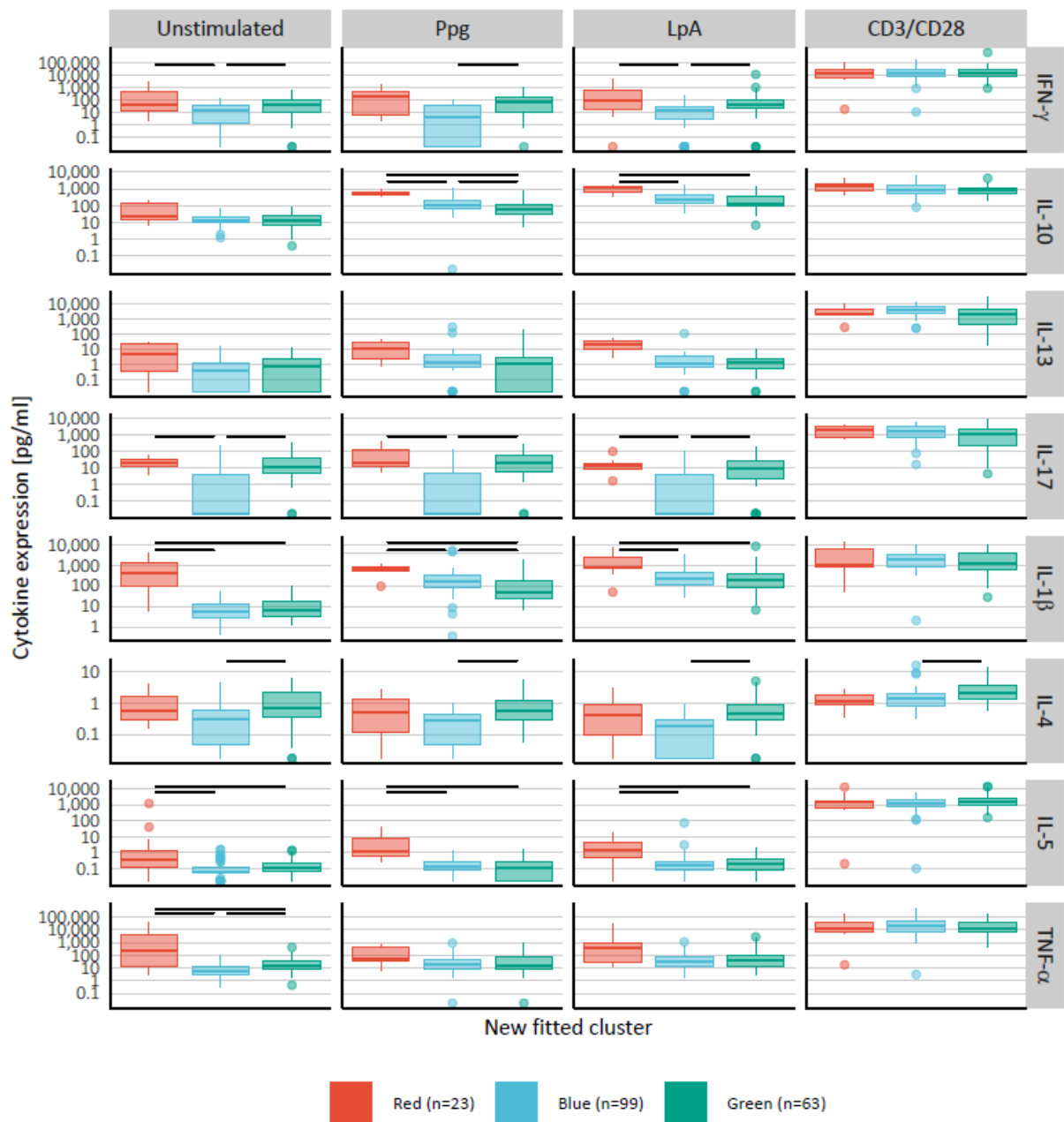
120

121 **Fig.S3**



122

123



125

126

