A Novel, Noninvasive Method to Diagnose Active Eosinophilic Esophagitis, Combining Clinical Data and Oral Cavity RNA Levels

Eosinophilic esophagitis (EoE) is a chronic inflammatory disease characterized by esophageal dysfunction and infiltration of eosinophils in the esophageal mucosa. According to the last published guideline for EoE diagnosis, obtaining at least 6 biopsies from a minimum of 2 esophageal locations is recommended, focusing on areas with endoscopic mucosal abnormalities. The cutoff of 15 eosinophils/per high-power field (HPF) is considered as the diagnostic threshold in a patient with symptoms of esophageal dysfunction.[1] Despite being the most reliable method for EoE diagnosis, endoscopic biopsies are an invasive procedure, so a noninvasive diagnostic test would be a great step forward for patients, especially for children.

Biomarkers based on disease-specific gene expression profiles are useful tools for disease prediction. EoE differential transcriptome analyses have been carried out,[2–5] and a predictive model based on gene expression profiles of esophageal biopsies has been recently developed.[3] Moreover, a non–biopsy-based predictive model for EoE diagnosis considering dysphagia and clinical data such as age, sex, and the presence of atopies has also been proposed.[6] However, the use of endoscopy and multiple biopsies cannot be discarded using these methods.

The oral mucosa is located close to the esophagus, and it has been suggested that it shares the same squamous epithelial pathology, containing many proteins and RNA molecules that could serve as biomarkers.[7] Considering that oral cavity samples are much more accessible than the esophageal tissue, our aim was to develop an alternative noninvasive approach for EoE diagnosis that would avoid biopsy acquisition. A schematic flowchart of the steps carried out to develop this approach is shown in Supplementary Figure 1*A*.

For the selection of potential biomarkers, we reanalyzed the most recent publicly available RNA sequencing (RNAseq) data from esophageal biopsies of controls and untreated, active EoE patients.[5] We analyzed the expression of the RNAseq candidate genes in the oral cavity samples from 10 EoE patients and 10 healthy donors (Testing cohort 1, Supplementary Table 1) to assess their utility in such samples. Only 43% of the genes analyzed (29 of 68) were expressed in oral samples (Supplementary Figure 1*B*), and thus, these were selected for differential expression analyses in a larger number of samples (Testing cohort 2, Supplementary Table 1).

Subsequently, differential expression analyses of the 29 expressed genes were carried out by reverse transcription quantitative polymerase chain reaction (qPCR) in oral samples of 19 untreated active EoE patients, 8 EoE patients presenting symptomatic relief as well as histologic remission with less than 15 eosinophils per HPF, considered as patients in clinical remission, and 13 controls (Testing cohort 2, Supplementary Table 1). Twenty-seven of the analyzed genes were protein-coding genes, and 2 were long noncoding RNAs (lncRNAs). *GAPDH* was used as endogenous housekeeping gene. The 8 genes with most significant alterations were selected for further potential biomarker validation (Figure 1*A*).

For the validation phase, the expression of 6 coding genes and 2 lncRNAs was evaluated in oral samples from 50 untreated active EoE patients, 24 EoE patients in clinical remission, and 40 controls (Validation cohort, cohort 3, Supplementary Table 1) (Figure 1*B*, Supplementary Figure 1*C* and *D*). *CDH26*, *KCNJ2,* and *PLD1* showed significant differential expression between controls and active EoE patients (Figure 1*B*). No differential expression was observed in any of these 8 genes when active EoE patients were compared with patients in clinical remission (Supplementary Figure 1*D*). Thus, subsequent analyses were performed by only using active EoE patients and controls.

Individual evaluation of the diagnostic performance of the 3 significant genes using receiver operating characteristic (ROC) analysis yielded an area under the curve (AUC) value of 0.72 for *CDH26* ($P = .0022$), 0.69 for *KCNJ2* ($P = .0059$), and 0.64 for *PLD1* ($P = .036$) (Figure 1*C*). The diagnostic performance of the combination of those genes was also evaluated. Combined values were calculated by logistic regression formulas to obtain probability values (*P* values), and combined ROC curve analysis was used to confirm discriminatory potential. These combined *P* values showed significant differences between active EoE patients and controls (Figure 1*D*). The combination of *CDH26*, *KCNJ2,* and *PLD1* yields a ROC-plot AUC of 0.78 ($P = .0002$) (Figure 1*E*), and the predictive model generated achieved 80.5% sensitivity and 66.7% specificity.

Most current article

1542-3565
https://doi.org/10.1016/j.cgh.2023.07.023

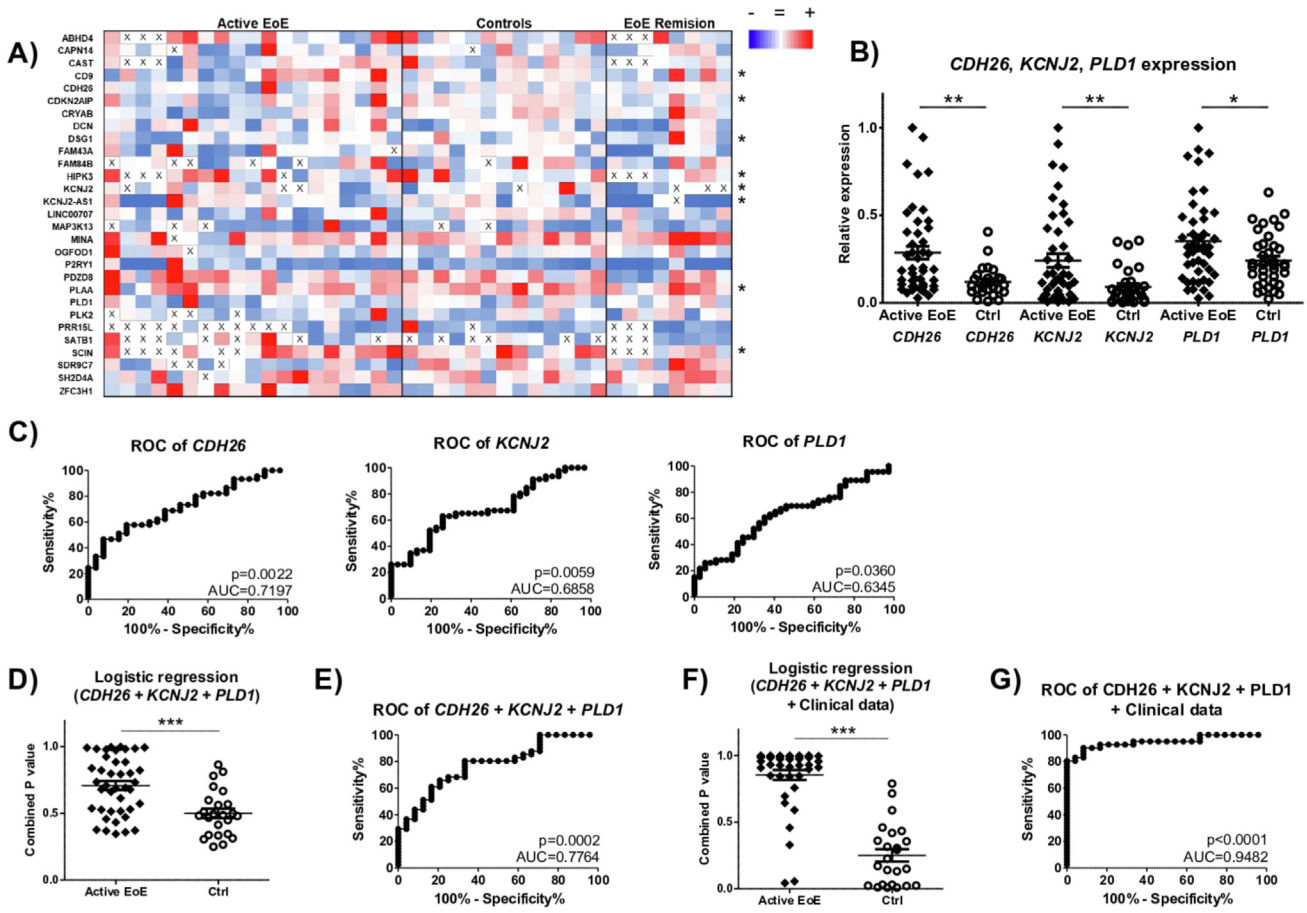


Figure 1. (*A*) Expression level heat map of the genes analyzed in testing cohort 2 (n = 40). Up-regulation is shown in *red* and down-regulation in *blue*. The 8 genes selected for further analysis are marked with an asterisk (*) on the right. (*B*) Differential expression of *CDH26*, *KCNJ2,* and *PLD1* genes between active EoE patients (n = 50) and controls (n = 40). (*C*) Individual ROC curves of *CDH26*, *KCNJ2,* and *PLD1* (n = 90). Combined *P* values of genes *CDH26*, *KCNJ2,* and *PLD1* in (*D*) and in addition to clinical data in (*F*). ROC-plot of the combined prediction model based on gene expression only in (*E*) and including clinical data in (*G*). *$P < .05$, **$P < .01$, ***$P < .001$ by Student *t* test.

In addition, adding clinical data such as age, sex, and the presence of related atopies significantly improved the predictive model. We calculated the new combined *P* values (Figure 1*F*) and generated an improved prediction model comprising oral biomarker expression and clinical data. This new combination yields a ROC-plot AUC of 0.95 ($P < .0001$) (Figure 1*G*) with 90.2% sensitivity and 91.7% specificity. Additional analyses confirmed that neither the sex nor the presence or the number of atopies significantly influenced gene expression.

Here we present a completely noninvasive approach useful for EoE diagnosis, which is based on objective and easy-to-collect clinical data and the expression of *CDH26*, *KCNJ2,* and *PLD1* in the oral cavity that can be easily assessed using commercial reagents. The information for the material and methods used can be found in the Supplementary Methods section. This combination efficiently works for initial diagnosis because it can distinguish active EoE patients from controls. Indeed, even if a first endoscopy might be needed to solve dysphagia, refractory chest pain, abdominal pain, vomiting, or heartburn symptoms, our approach would avoid a subsequent esophageal biopsy acquisition usually performed to confirm or discard EoE. Other combinations will need to be explored for the assessment of patients on EoE clinical remission or suffering from other esophageal diseases such as gastroesophageal reflux disease, opening the door to an interesting set of future studies.

Our prediction model showed high sensitivity and specificity scores, which is similar to the previously proposed EoE diagnostic panel,[3] which is based on the gene expression profile of esophageal biopsies. However, the use of oral cavity samples instead of esophageal biopsies enables an easier and much more comfortable procedure, in addition to being a faster and more cost-effective way to diagnose EoE.

MAIALEN SEBASTIAN-DELACRUZ
Department of Genetics, Physical Anthropology and Animal Physiology
University of the Basque Country (UPV/EHU)
Leioa, Spain, *and*

Biocruces Bizkaia Health Research Institute
Barakaldo, Spain

KOLDO GARCIA-ETXEBARRIA
Department of Gastroenterology
Biodonostia Health Research Institute
Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd)
University of the Basque Country (UPV/EHU)
San Sebastián, Spain

JOSE RAMÓN BILBAO
Department of Genetics
Physical Anthropology and Animal Physiology
University of the Basque Country (UPV/EHU)
Leioa, Spain, *and*
Biocruces Bizkaia Health Research Institute,
Barakaldo, Spain, *and*
CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM)
Instituto de Salud Carlos III
Madrid, Spain

ALFREDO J. LUCENDO
Department of Gastroenterology
Hospital General de Tomelloso
Tomelloso, Spain, *and*
CIBER de enfermedades Hepáticas y Digestiva (CIBERehd)
Instituto de Investigación Sanitaria La Princesa
Instituto de Investigación Sanitaria de Castilla-La Mancha (IDISCAM)

LUIS BUJANDA
Department of Gastroenterology
Biodonostia Health Research Institute
Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd)
University of the Basque Country (UPV/EHU)
San Sebastián, Spain

AINARA CASTELLANOS-RUBIO
Department of Genetics, Physical Anthropology and Animal Physiology
University of the Basque Country (UPV/EHU)
Leioa, Spain, *and*
Biocruces Bizkaia Health Research Institute,
Barakaldo, Spain, *and*
CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM)
Instituto de Salud Carlos III
Madrid, Spain, *and*
Ikerbasque, Basque Foundation for Science
Bilbao, Spain

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at https://doi.org/10.1016/j.cgh.2023.07.023.

References

1. Lucendo AJ, et al. United Eur Gastroenterol J 2017;5:335–358.
2. Blanchard C, et al. J Clin Investig 2006;116:536–547.
3. Wen T, et al. Gastroenterology 2013;145:1289–1299.
4. Sherrill JD, et al. Genes Immun 2014;15:361–369.
5. Wheeler JC, et al. J Allergy Clin Immunol 2019;143:2131–2146.
6. Cotton CC, et al. Clin Gastroenterol Hepatol 2021;19:1824–1834.
7. Hayat JO, et al. Gut 2015;64:373–380.

Correspondence
Address correspondence to: Ainara Castellanos-Rubio, PhD, University of the Basque Country (UPV/EHU), Sarriena s/n, 48940 Leioa, Spain. e-mail: ainara.castellanos@ehu.eus.

Conflicts of interest
The authors disclose no conflicts.

Funding
Supported by a grant from the Spanish Ministry of Science, (SAF2017-91873-EXP), a grant from the Department of Health from the Basque Government (EJ-2017111082), a research fellowship from the Asociación de Celiacos y Sensibles al Gluten de Madrid (A.C.R.), a grant from the Spanish Ministry of Science and Innovation (PID2019-106382RB-I00) (JRB), and the Asociación de Investigación Biomédica La Mancha Centro (AJL). MSdlC was supported by predoctoral fellowship from the Universiti Teknologi Brunei Country.

Supplementary Methods

Patients and Oral Cavity Sample Collection

Oral cavity samples were collected using swabs (Zymo Research, Irvine, CA, #R1107-E) at Biodonostia Health Research Institute and Hospital General de Tomelloso (Spain) from patients with clinical and histologically confirmed EoE, diagnosed following accepted criteria.[1] Healthy volunteers were included as controls. The study was approved by the corresponding Ethics Committees (BUJ-BIO-2020-01 and 137-C). Informed consent was obtained from all participants. Patients' clinical features are summarized in Supplementary Table 1.

Candidate Gene Selection

Normalized gene expression data were retrieved from RNAseq study GSE113341 deposited in Gene Expression Omnibus database.[5] Normalized expression levels of 25,474 genes were compared for differential expression using Mann-Whitney *U* test, and the significant genes, after false discovery rate test correction, were used for experimental biomarker validation.

RNA Extraction and Gene Expression Analysis

Total RNA was extracted using Trizol (Zymo Research, #R2053). Fifty to 150 ng of RNA was used for reverse transcription qPCR using PrimeTime One-Step RT-qPCR Master Mix (IDT, Inc, Coralville, IA, #10007067) and predesigned qPCR Assays for RNA quantitation (IDT, Inc). Reactions were run in duplicate in a BioRad CFX384 real-time PCR instrument. Expression levels were analyzed using the 2−ΔΔCt method. The housekeeping genes analyzed were *GAPDH*, *RPLP0,* and *18S*, and the least variable was selected for relative expression normalization. All the candidate genes analyzed and used assays and their sequences are available upon request.

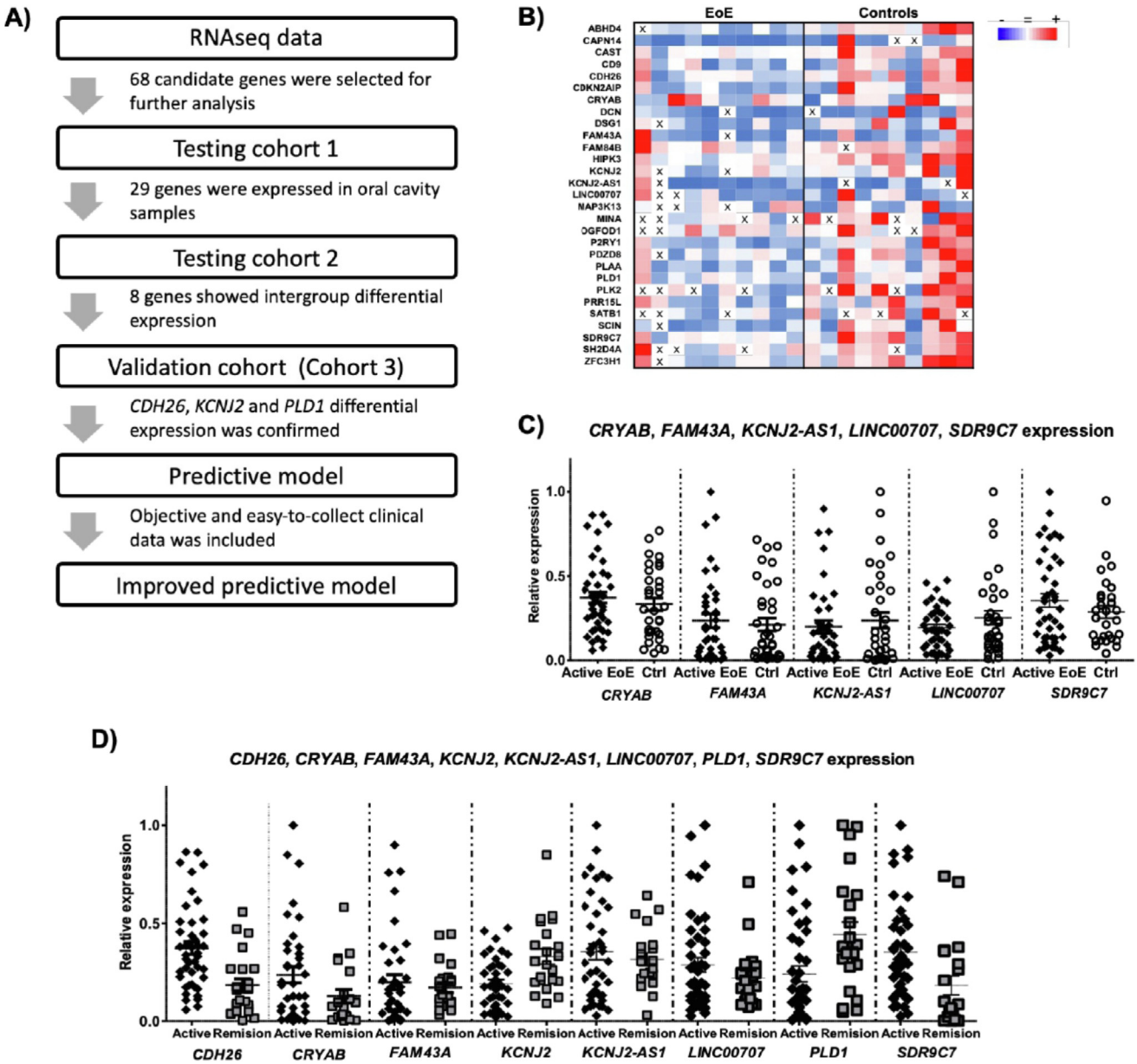
Expression of the candidate and housekeeping genes was first quantified in oral cavity samples from 10 EoE patients and 10 controls (cohort 1). Genes for which amplification was observed in more than half of the samples were selected for further analyses in a larger cohort of active EoE patients (n = 19), controls (n = 13), and EoE patients in clinical remission (n = 8) (cohort 2). Genes with differential expression among groups were analyzed in a confirmatory cohort of EoE active patients (n = 50), non-EoE controls (n = 40), and EoE patients in clinical remission (n = 24) (cohort 3). Those genes that showed differential expression between active EoE patients and controls were evaluated as possible noninvasive biomarkers for EoE diagnosis by calculating the ROC curves.

Predictive Model Generation

Combinatorial analysis of the significant genes was performed by using logistic regression formulas. Combined *P* values were obtained by an online logistic regression calculator (AAT Bioquest, Inc) and were used for combined ROC curve analysis. Clinical data added to the model were first given a numeric value when needed. Age values were considered the exact age when the oral cavity samples were collected; sex punctuation was divided into 1 for male and 0 for female; and in the field of related atopies a value of 1 was given if patients had asthma, rhinitis, conjunctivitis, dermatitis, and food and other kind allergies, and a 0 was given if not.

Statistical Analysis

Differential expression was considered significant when $P < .05$ using Student *t* test. Outlier values were discarded by using ROUT method with a Q of 1%.



Supplementary Figure 1. (*A*) Schematic flowchart of the steps followed for the development of the prediction model for EoE diagnosis. (*B*) Heat map of the expression of candidate genes identified in the RNAseq data set expressed in testing cohort 1 (n = 20). Up-regulation is shown in *red* and down-regulation in *blue*. Comparison of gene expression in the validation cohort (cohort 3), between active EoE patients (n = 50) and controls (n = 40) (*C*), and between active EoE patients (n = 50) and EoE patients with clinical remission (n = 24) (*D*).

Supplementary Table 1. Clinical Characteristics of Patients Included in Our Study

Cohort	Diagnosis	Eosinophils counts	Age	Sex	Related atopies	Kind of atopy
Testing cohort 1 (n = 20)	EoE (n = 10)	50 ± 22	41 ± 15	Male 80.0% Female 20.0%	At least one atopy 40.0% More than one atopy 30.0% No atopies 60.0%	Asthma 10.0% Rhinitis 20.0% Conjunctivitis 10.0% Dermatitis 10.0% Food allergies 10.0% Non-food allergies 20.0%
	Controls (n = 10)	N/A	37 ± 14	Male 30.0% Female 70.0%	No atopies 100.0%	
Testing cohort 2 (n = 40)	Active EoE (n = 19)	61 ± 37	33 ± 13	Male 78.9% Female 21.1%	At least one atopy 78.9% More than one atopy 52.6% No atopies 21.1%	Asthma 26.3% Rhinitis 57.9% Conjunctivitis 52.6% Dermatitis 5.3% Food allergies 42.8%
	Controls (n = 13)	0	34 ± 16	Male 38.5% Female 61.5%	At least one atopy 7.7% More than one atopy 7.7% No atopies 92.3%	Asthma 15.4% Rhinitis 7.7%
	EoE in remission (n = 8)	6 ± 5	37 ± 14	Male 62.5% Female 37.5%	At least one atopy 37.5% More than one atopy 12.5% No atopies 60.5%	Asthma 25.0% Rhinitis 25.0% Conjunctivitis 12.5% Food allergies 12.5%
Validation cohort (cohort 3) (n = 114)	Active EoE (n = 50)	52 ± 33	31 ± 14	Male 80.0% Female 20.0%	At least one atopy 80.0% More than one atopy 60.0% No atopies 20.0%	Asthma 38.0% Rhinitis 52.0% Conjunctivitis 40.0% Dermatitis 20.0% Food allergies 30.0% Non-food allergies 10.0%
	Controls (n = 40)	0	47 ± 13	Male 47.5% Female 52.5%	At least one atopy 22.5% More than one atopy 7.5% No atopies 77.5%	Asthma 7.5% Rhinitis 5.0% Conjunctivitis 2.5% Dermatitis 2.5% Non-food allergies 12.5%
	EoE in remission (n = 28)	6 ± 6	34 ± 15	Male 70.8% Female 29.2%	At least one atopy 87.5% More than one atopy 70.8% No atopies 12.5%	Asthma 41.7% Rhinitis 54.2% Conjunctivitis 58.3% Dermatitis 16.7% Food allergies 33.3% Non-food allergies 12.5%

NOTE. Mean and standard deviation of number of cells/HPF and years are shown for eosinophils counts and age, respectively.