

Inflammatory endotypes of chronic rhinosinusitis based on cluster analysis of biomarkers

Peter Tomassen, MD,^a Griet Vandeplass, MD,^a Thibaut Van Zele, MD, PhD,^a Lars-Olaf Cardell, MD, PhD,^b Julia Arebro, MD,^b Heidi Olze, MD, PhD,^c Ulrike Förster-Ruhrmann, MD,^c Marek L. Kowalski, MD, PhD,^d Agnieszka Olszewska-Zięber, MD,^d Gabriele Holtappels,^a Natalie De Ruyck,^a Xiangdong Wang, MD, PhD,^e Cornelis Van Drunen, PhD,^f Joaquim Mullol, MD, PhD,^g Peter Hellings, MD, PhD,^h Valerie Hox, MD, PhD,^h Elina Toskala, MD, PhD,ⁱ Glenis Scadding, MD,^j Valerie Lund, MD,^j Luo Zhang, MD, PhD,^e Wytse Fokkens, MD, PhD,^f and Claus Bachert, MD, PhD^{a,b}
Ghent and Leuven, Belgium, Stockholm, Sweden, Berlin, Germany, Lodz, Poland, Beijing, China, Amsterdam, The Netherlands, Barcelona, Spain, Philadelphia, Pa, and London, United Kingdom

Background: Current phenotyping of chronic rhinosinusitis (CRS) into chronic rhinosinusitis with nasal polyps (CRSwNP) and chronic rhinosinusitis without nasal polyps (CRSSNP) might not adequately reflect the pathophysiologic diversity within patients with CRS.

Objective: We sought to identify inflammatory endotypes of CRS. Therefore we aimed to cluster patients with CRS based solely on immune markers in a phenotype-free approach.

Secondarily, we aimed to match clusters to phenotypes.

Methods: In this multicenter case-control study patients with CRS and control subjects underwent surgery, and tissue was analyzed for IL-5, IFN- γ , IL-17A, TNF- α , IL-22, IL-1 β , IL-6,

IL-8, eosinophilic cationic protein, myeloperoxidase, TGF- β 1, IgE, *Staphylococcus aureus* enterotoxin-specific IgE, and albumin. We used partition-based clustering.

Results: Clustering of 173 cases resulted in 10 clusters, of which 4 clusters with low or undetectable IL-5, eosinophilic cationic protein, IgE, and albumin concentrations, and 6 clusters with high concentrations of those markers. The group of IL-5–negative clusters, 3 clusters clinically resembled a predominant chronic rhinosinusitis without nasal polyps (CRSSNP) phenotype without increased asthma prevalence, and 1 cluster had a T_H17 profile and had mixed CRSSNP/CRSwNP. The IL-5–positive clusters were divided into a group with moderate IL-5 concentrations, a mixed CRSSNP/CRSwNP and increased asthma phenotype, and a group with high IL-5 levels, an almost exclusive nasal polyp phenotype with strongly increased asthma prevalence. In the latter group, 2 clusters demonstrated the highest concentrations of IgE and asthma prevalence, with all samples expressing *Staphylococcus aureus* enterotoxin-specific IgE.

Conclusion: Distinct CRS clusters with diverse inflammatory mechanisms largely correlated with phenotypes and further differentiated them and provided a more accurate description of the inflammatory mechanisms involved than phenotype information only. (J Allergy Clin Immunol 2016;■■■:■■■-■■■.)

Key words: Chronic rhinosinusitis, nasal polyps, endotypes, cluster analysis, inflammation, asthma

From ^athe Upper Airways Research Laboratory, Ghent University; ^bthe Division of ENT Diseases, CLINTEC, Karolinska Institutet, Stockholm; ^cthe Department of Otorhinolaryngology, Charité-Universitätsmedizin Berlin; ^dthe Department of Immunology, Rheumatology and Allergy, Medical University of Lodz; ^ethe Department of Otolaryngology–Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing; ^fthe Department of Otorhinolaryngology, Academic Medical Centre, Amsterdam; ^gClinical and Experimental Respiratory Immunology, Hospital Clínic, IDIBAPS, Barcelona; ^hthe Department of Otorhinolaryngology, Head and Neck Surgery, University Hospitals Leuven; ⁱthe Department of Otolaryngology–Head and Neck Surgery, Temple University, School of Medicine, Philadelphia; and ^jRoyal National Throat, Nose and Ear Hospital, London.

Supported by the Sixth European Union Framework Program for Research (contract no. FOOD-CT-2004-506378) and Fonds Wetenschappelijk Onderzoek Vlaanderen, International Coordination Action (no. G.0854.09).

Disclosure of potential conflict of interest: P. Tomassen has received research support from the Sixth European Union Framework Program for Research and Fonds Wetenschappelijk Onderzoek Vlaanderen. C. Van Drunen has received research support from the European Union FP6 (GA2LEN Network of Excellence), European Union (BM4SIT), GlaxoSmithKline, ALK-Abelló, and Allergopharma. J. Mullol is on the boards for Uriach, Meda, Johnson & Johnson, FAES, Crucell, and ALK-Abelló; has received research support from GlaxoSmithKline, Iriach, FAES, and Meda; and has received lecture fees from Uriach, Hartington Pharmaceuticals, Novartis, FAES, Menarini, MSD, and Pierre-Fabre. E. Toskala has received consultancy fees from Merck as an Advisory Board member. G. Scadding has received consultancy fees from ALK-Abelló, Meda, and GlaxoSmithKline; has received research support from GlaxoSmithKline; and has received lecture fees from ALK-Abelló and Meda. V. Lund has received research support from GlaxoSmithKline and Optinose; has received lecture fees from MSD; and receives royalties as Editor of *Cummings Otorhinolaryngology*, Elsevier. W. Fokkens has received research support from Meda and has received payment for developing educational presentation (webcast on treatment of rhinitis for general practitioners). The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication March 23, 2015; revised December 13, 2015; accepted for publication December 21, 2015.

Corresponding author: Claus Bachert, MD, PhD, Upper Airway Research Laboratory, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium. E-mail: claus.bachert@ugent.be.

0091-6749/\$36.00

© 2016 American Academy of Allergy, Asthma & Immunology

<http://dx.doi.org/10.1016/j.jaci.2015.12.1324>

Chronic rhinosinusitis (CRS) is a disabling disease affecting 10.9% of the European¹ and 13.4% of the American² general population. CRS is defined by symptoms and clinical signs, and diagnosis can be supported by nasal endoscopy and computed tomographic (CT) scanning.³ In a European multicenter epidemiologic study involving 56,000 subjects, it has been demonstrated that cigarette smoking increases the risk of CRS¹; of interest, CRS is also associated with late-onset asthma.⁴ However, questionnaire-based population studies are limited in further defining possible subgroups relevant for these associations.⁵ Indeed, CRS shows remarkable heterogeneity, both at the clinical phenotype level and at the molecular pathophysiologic level. Current consensus in Europe and the United States discerns 2 major phenotypes, which were defined as subgroups of patients with homogeneous clinically observable characteristics⁶ based on nasal endoscopic findings: chronic rhinosinusitis with nasal polyps (CRSwNP) and chronic rhinosinusitis without nasal polyps (CRSSNP). Furthermore, there are additional subtypes,

Abbreviations used

CRS: Chronic rhinosinusitis
 CRSsNP: Chronic rhinosinusitis without nasal polyps
 CRSwNP: Chronic rhinosinusitis with nasal polyps
 ECP: Eosinophilic cationic protein
 MPO: Myeloperoxidase
 SE-IgE: *Staphylococcus aureus* enterotoxin-specific IgE

such as allergic fungal rhinosinusitis, CRS associated with aspirin-exacerbated respiratory disease, and CRS associated with cystic fibrosis, which can present as CRSwNP or CRSsNP.³

The clinical dichotomization of CRSwNP versus CRSsNP was initially reflected at the molecular level, with a predominance of T_H1 cells in patients with CRSsNP and T_H2 cells and eosinophils in patients with CRSwNP among white subjects.^{7,8} However, subsequent studies reported a wider spectrum of immunologic profiles, especially in nonwhite phenotypes,⁹ expressing a neutrophilic type of inflammation with involvement of other T-cell subsets, such as T_H1 and T_H17 cells. Furthermore, the simultaneous expression of different T_H cell types within a single tissue was demonstrated.¹⁰

Thus the simple differentiation in T_H1 and T_H2 disease does not encompass the molecular diversity in patients with CRS, and the clinical phenotype does not adequately identify the immunologic profile. Therefore we sought to identify inflammatory endotypes of CRS, which were defined as “subtypes of disease with a unique pathomechanism, functionally and pathologically different from others by the involvement of a specific molecule or cell.”¹¹ For this, we aimed to cluster patients with CRS solely based on immune markers in a clinical phenotype-free approach. Secondly, we aimed to match these endotypes with clinical phenotypes and selected clinical parameters, of which we already have identified an association with immune markers.^{12,13} This approach would also facilitate the identification of therapeutic targets and predict response to those approaches, such as biologicals.

METHODS**Study design**

The study was designed as a multicenter case-control study carried out by the GA²LEN Sinusitis Cohort group (principal investigator, C. Bachert) in the framework of the European FP6 research initiative. The ear, nose, and throat departments in tertiary referral academic hospitals of Ghent, Leuven, Amsterdam, Barcelona, London, Berlin, Helsinki, Lodz, Malmö, and Stockholm participated in this study. This study was approved by the ethics committees of all individual institutions involved in data and tissue collection. Informed consent was obtained from all subjects before sample collection.

Patient population

Patients with CRS (defined by EPOS criteria [European Position Paper on Rhinosinusitis and Nasal Polyps] based on symptoms and results of nasal endoscopy and computed tomography of the sinuses³) were included as cases, and subjects undergoing inferior turbinate surgery with no signs of CRS were included as control subjects. Excluded were subjects with an acute exacerbation of rhinosinusitis 2 weeks before inclusion, subjects with immunodeficiencies, subjects with cystic fibrosis, and subjects who used oral or nasal steroids in the 4 weeks preceding surgery or antileukotrienes in the 2 weeks before inclusion. In total, 917 subjects were recruited, of whom there were 682 cases (65% with CRSwNP) and 187 control subjects. Subjects underwent a standardized skin prick test¹⁴ and were registered as asthmatic

based on a clinical diagnosis. Patients were considered allergic on clinical grounds (positive skin prick test or provocation test results or specific IgE levels to inhalant allergens plus allergy symptoms present).

At a maximum of 12 months after inclusion, a portion of the subjects underwent functional endoscopic sinus surgery, and control subjects underwent partial inferior turbinotomy during septal surgery (226 patients with CRS and 106 control subjects). Control subjects with allergic rhinitis were not excluded. The indications for surgery and its procedures were based on clinical decisions independent from participation in the study. Tissue was collected from nasal polyps or sinus mucosa in patients with CRS and from the inferior turbinates in control subjects. Tissue was separated from bone fragments, snap-frozen in liquid nitrogen, and stored at -80°C . Control tissues were not used in cluster analyses of the patients with CRS but solely served to determine normal values.

Measurement of inflammatory markers

For tissue analysis, we selected 14 markers (Table 1) based on earlier publications; those markers reflect the inflammatory patterns observed in patients with CRSsNP and patients with CRSwNP. Only subjects for whom there was an adequate amount of tissue that was needed for these analyses (typically 0.15 g) were included. Briefly, as described before,¹⁵ each 0.1 g of tissue was diluted in 1 mL of 0.9% NaCl solution containing a protease inhibitor cocktail (Complete; Roche Diagnostics, Mannheim, Germany), homogenized at 1000 rpm for 5 minutes, and centrifuged at 1500g for 10 minutes at 4°C . Concentrations of eosinophilic cationic protein (ECP), total IgE, and IgE specific to a mixture of *S aureus* enterotoxins (staphylococcal enterotoxins A and C and toxic shock syndrome toxin 1) were assayed by using the UniCAP system (Phadia, Uppsala, Sweden). Concentrations of IL-22, IFN- γ , and TGF- β 1 were assayed by using commercially available ELISA kits from R&D systems (Minneapolis, Minn). Myeloperoxidase (MPO) concentrations were measured with a commercially available ELISA from BioCheck (Foster City, Calif). For albumin, we used kits from AssayPro (St Charles, Mo). Concentrations of IL-1 β , IL-5, IL-6, IL-8, IL-17A, and TNF- α were assayed with the Luminex 100 system (Luminex, Austin, Tex).

Concentrations in tissue homogenates were expressed as mass versus volume after multiplication by a homogenization dilution factor of 11. Values of less than the limit of detection were considered negative for categorical analysis and given a value equal to half of the detection limit for continuous analysis.

Statistical methods

R version 2.15.2 statistical software was used.¹⁶ The following variables, which had a high proportion (>33%) of values of less than the detection limit, were used as binomial variables, with their detection limits used as a cutoff: *Staphylococcus aureus* enterotoxin-specific IgE (SE-IgE; cutoff, 3.85 kU_A/L), IFN- γ (cutoff, 85.8 pg/mL), TNF- α (cutoff, 38.94 pg/mL), IL-17A (cutoff, 25.06 pg/mL), and IL-5 (cutoff, 12.98 pg/mL). Baseline parameters were tested for differences between cases and control subjects by using t tests on continuous variables and χ^2 tests on discrete variables.

To analyze relationships between variables, we performed principal component analysis, after which orthogonal rotation with Kaiser normalization was performed, and only variables with loadings of greater than 0.4 were retained. Next, we performed cluster analysis on the variables by using ascendant hierarchical clustering based on the correlation ratio and mixed principal component analysis,¹⁷ and the optimal number of clusters was determined by using the Rand statistic.

For cluster analysis of individual cases, a Gower dissimilarity matrix was calculated, which allows both ordinal and binomial data. Next, clusters were calculated by using the Partitioning Around Medoids method, with cluster numbers ranging from 2 to 15 generated. The optimal number of clusters was based on the elbow (maximum change) of the scree plot of the mean silhouette width, the Baker-Hubert Gamma statistic,¹⁸ and the Hubert-Levin C index.¹⁹ Additionally, the optimal number of clusters was assessed by using Jaccard stability after bootstrap resampling²⁰ and visual inspection of the clusplot after multidimensional scaling.²⁰

TABLE I. Markers used for component and cluster analysis

Marker	Cutoff value	Interpretation of increased concentrations
IFN- γ	85.8 pg/mL	T _H 1 activity
IL-5	12.98 pg/mL	T _H 2 activity
IL-17A	25.06 pg/mL	T _H 17 activity
IL-22	NA	T _H 22 activity
TNF- α	38.94 pg/mL	Proinflammatory action
IL-1 β	NA	Proinflammatory action
IL-6	NA	Proinflammatory action
IL-8	NA	Neutrophilic chemotaxis
MPO	NA	Neutrophilic activity marker
ECP	NA	Eosinophilic activity marker
IgE	NA	Adaptive immunity marker
SE-IgE	3.85 kU _A /L	Marker for superantigen effect on local mucosa
Albumin	NA	Edema
TGF- β 1	NA	Fibrosis/regulatory T-cell activation

Cutoff values are given for markers that were analyzed as categorical, and a short description of the interpretation was given regarding the underlying inflammatory pattern when an increased concentration or an increased proportion above the cutoff value was observed. Cutoff values are expressed as mass per volume of undiluted homogenized tissue (ie, calculated as the detection limit $\times$ dilution factor).

For descriptive characterization of individual clusters, cytokine and phenotype data from each cluster were tested for difference from the control group (Mann-Whitney *U* test and logistic regression). Next, between-cluster differences of all parameters, as well as phenotype parameters, were tested by using the Kruskal-Wallis test with multiple group comparisons and Benjamini-Hochberg adjustment for multiple testing or Tukey contrasts in logistic models for binomial parameters. For visualization of clusters, a “clusplot” was drawn, plotting subjects in 2 dimensions after multidimensional scaling.

As an aid in interpretation of inflammatory patterns, levels of each cytokine in each cluster were categorized as increased compared with control values only, increased compared with control values plus 2, 3, or 6 other clusters, or categorized as no difference from control values or other clusters. Furthermore, for each cluster, we interpreted inflammatory patterns based on a high proportion of positive categorical variables or increased concentrations of continuous variables compared with control values and other clusters. The pattern for interpretation is listed in Table I. A modified heat map was created after ordering and grouping clusters with similar characteristics and ordering variables according to the previous principal component analysis. The variables that were used categorically in the cluster analysis were additionally tabulated with their continuous values.

RESULTS

Of the 226 cases and 106 control subjects who underwent surgery, 173 cases and 89 control subjects had an adequate amount of tissue collected to carry out all intended analyses. Demographic, phenotype, and cytokine data of cases and control subjects are tabulated in Table II; demographic data of the sample included were not different from those of the total cohort. Cases had a significantly higher age and a higher prevalence of smoking history and asthma. Cases had significantly higher concentrations than control subjects of IL-6, IL-8, albumin, MPO, ECP, and IgE and had a significantly higher proportion of concentrations of greater than the detection limit for IL-5, IFN- γ , and SE-IgE (Table II).

Principal component analysis (as shown in Fig 1, A and B, and see Table E1 in this article's Online Repository at www.jacionline.org) retained 5 components explaining 74% of all variance in the data. The first component was composed of MPO, IL-1 β , IL-6, and IL-8. The second component consisted of IgE, ECP, IL-5, and albumin. The third component included

TABLE II. Characteristics of cases and control subjects

	Cases	Control subjects	P value, cases vs control subjects
Mean age (y)	42.8	32.8	<.001*
Sex (% male)	58.2	58.1	.995*
Proportion ever smoker (%)	64.7	47.1	.005†
Proportion current smoker (%)	29.0	25.9	.592†
Proportion with allergy (%)	45.0	34.1	.084†
Proportion with asthma (%)	28.2	15.3	.019†
Proportion with aspirin sensitivity (%)	4.5	1.1	.148†
Proportion with CRSwNP (%)	56.6	—	—
Total IgE (kU/L, geometric mean)	87.6	8.9	<.001*
ECP (μ g/L, geometric mean)	2348.5	116.6	<.001*
IL-8 (pg/mL, geometric mean)	1798	644	<.001*
IL-6 (pg/mL, geometric mean)	72.8	41.0	.021*
IL-1 β (pg/mL, geometric mean)	28.6	23.7	.392*
TGF- β 1 (pg/mL, geometric mean)	9358	8248	.228*
MPO (ng/mL, geometric mean)	2122	966	<.001*
IL-22 (pg/mL, geometric mean)	363.2	347.5	.488*
Albumin (μ g/dL, geometric mean)	965.2	552.4	<.001*
IL-17A (% positivity)	32.1	25.3	.239†
TNF- α (% positivity)	58.4	48.3	.108†
SE-IgE (% positivity)	12.9	0.0	<.001†
IFN- γ (% positivity)	26.5	12.5	.015†
IL-5 (% positivity)	63.4	3.5	<.001†

Concentrations are expressed as mass per volume of undiluted homogenate.

*positivity, Proportion samples greater than the cutoff value.

†Student *t* test.

‡Pearson χ^2 test.

TNF- α , IL-17, and IL-22. The fourth component was composed of TGF- β and SE-IgE, and the last component was composed of IFN- γ and SE-IgE. In a hierarchical cluster analysis of variables (shown in Fig 1, C), these were optimally clustered in the 6 following clusters: (1) IgE, ECP, IL-5, albumin, and SE-IgE; (2) IL-1, IL-6, IL-8, and MPO; (3) IL-17 and TNF- α ; (4) TGF- β 1; (5) IL-22; and (6) IFN- γ .

Clustering of all patients with CRS based on cytokine measurements, irrespective of phenotype, resulted in an optimal outcome of 10 clusters (details are given in Fig E1 and Table E2 in this article's Online Repository at www.jacionline.org). The clusters were well separated from each other, as shown in the cluster plot (Fig 2). Tests for between-cluster differences and differences from control subjects showed that for all cytokines, except TGF- β 1 and IL-1 β , there was at least 1 cluster with significantly higher concentrations than any other cluster. Compared with the control group, levels of all cytokines, except TGF- β 1, were significantly increased ($P < .05$) in at least 1 cluster. Not all clusters with a significant difference from the control group showed differences with other clusters. IL-22 concentrations did not differ significantly from control concentrations; however, there were multiple significant between-group differences.

Analogously, clinical parameters (which were not used in the cluster analysis) were analyzed within clusters and compared with control values (Fig 3). The proportion of CRSwNP was significantly different between clusters, with clusters 1 to 3 having a significantly lower proportion than most clusters and clusters 8 to 10 having a significantly higher proportion than most clusters. Asthma prevalence was significantly different

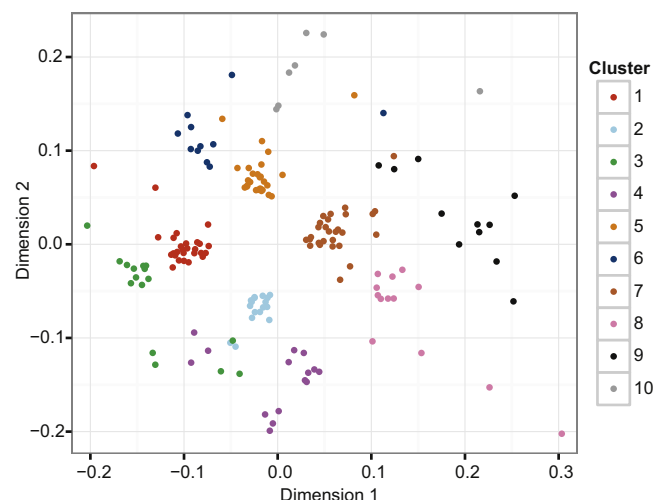
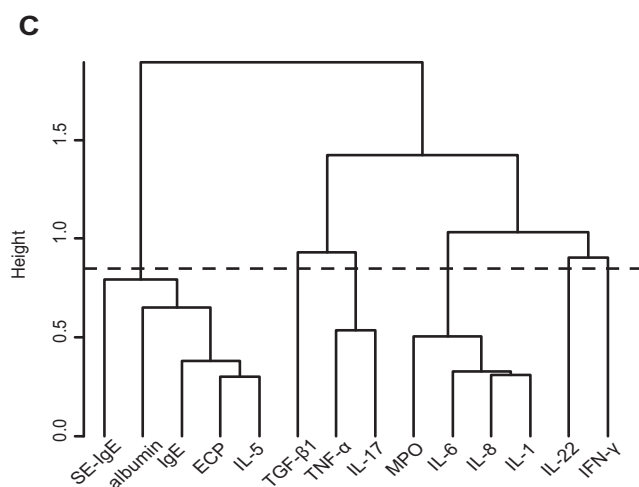
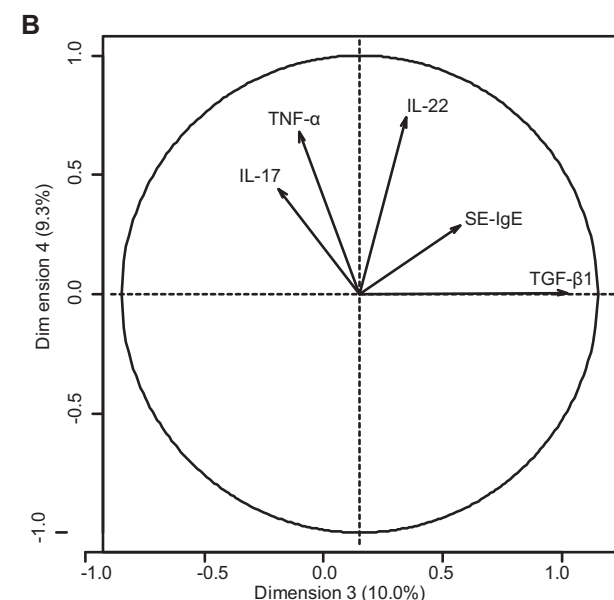
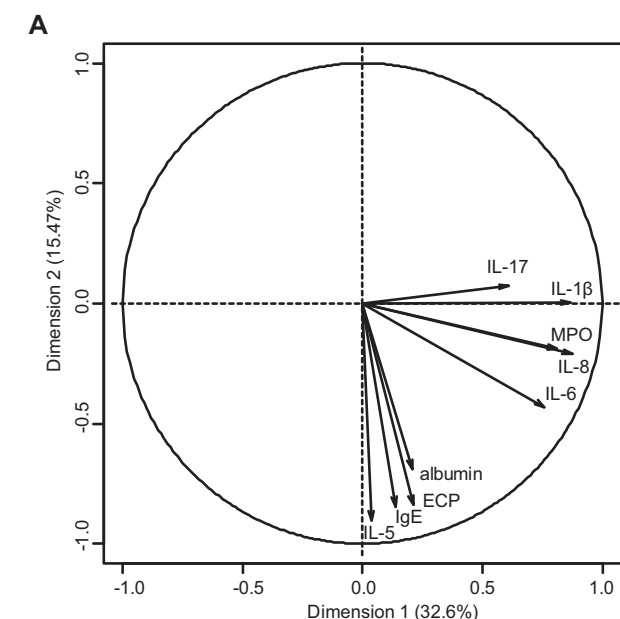


FIG 2. Cluster plot of cases. Clustering of cases illustrated in a clusplot in which patients are plotted in a 2-dimensional space after multidimensional scaling. Here, dissimilarities between patients, taking all variables into account, are represented by distances between the patients in 2 dimensions.

from that in control subjects and differed significantly between clusters. Age, sex, and prevalence of allergy and reported aspirin sensitivity were not significantly different between clusters (see Table E3 in this article's Online Repository at www.jacionline.org).

Means and ratios of cytokine and phenotype data were calculated and tabulated as a heat map to characterize the clusters (Fig 3). The most striking differentiation was formed by the IL-5/ECP/IgE/albumin axis, resulting in 4 clusters with undetectable or low concentrations of these biomarkers and 6 clusters with high concentrations. The endotype differentiation was also reflected in the clinical phenotypes, with 3 clusters (clusters 1-3) showing a predominant CRSsNP appearance, 4 clusters (clusters 4-7) showing a mixed phenotype, and 3 clusters (clusters 8-10) showing a predominant CRSwNP appearance. Clusters 1 to 3 showed no increase in asthma prevalence, whereas clusters 4 to 10 showed that asthma prevalence was increased compared with that in control subjects and clusters 1 to 3, with further increases in clusters 9 and 10.

The 4 clusters with predominant noneosinophilic appearance could further be differentiated as follows: cluster 1 did not show any signs of inflammation (levels of no single cytokine were significantly increased compared with control values in this cluster); cluster 2, a pure CRSsNP cluster, overexpressed TNF- α and IL-22; and cluster 3 overexpressed IFN- γ and MPO/IL-8. Cluster 4, in contrast, showed neutrophilic inflammation with upregulation of the proinflammatory cytokines IL-1 β /IL-6, IL-8/MPO, and IL-17A, IL-22, TNF- α , and partially IFN- γ .

FIG 1. Principal component analysis and clustering of variables. Vectors of each parameter are plotted against the first 4 rotated principal components. **A**, First 2 principal components. **B**, Third and fourth principal components. The fifth rotated component is not depicted. It consisted of IFN- γ and SE-IgE. **C**, Dendrogram of clustered variables. The horizontal dashed line indicates the tree cut used in this analysis based on the Rand index, resulting in 6 variable clusters.

Cluster																											
	N	N/total		MPO (ng/mL)	IL-8 (pg/mL)	IL-6 (pg/mL)	IL-1 β (pg/mL)	Albumin (mg/dL)	IL-5 (pg/mL)	IL-5 (ratio pos.)	ECP (μ g/L)	IgE (kU/L)	SE-IgE (kU/L)	SE-IgE (ratio pos.)	TGF- β 1 (pg/mL)	IL-17 (pg/mL)	IL-17 (ratio pos.)	TNF- α (pg/mL)	TNF- α (ratio pos.)	IL-22 (pg/mL)	IFN- γ (pg/mL)	IFN- γ (ratio pos.)	prop. CRSwNP	prop. asthma	prop. allergy		
1	27	16%		973	956	30	24	65	1	0%	407	20	1.29	7%	17853	0.1	0%	0.8	0%	307	43	0%	11%	15%	42%		
2	17	10%		1299	442	5	14	54	1	0%	284	13	1.01	0%	13676	0.1	0%	5.4	100%	463*	48	12%	0%	7%	63%		
3	16	9%		2714	2908	91	69	70	2	0%	498	28	1.18	6%	17111	0.1	13%	1.2	19%	311	213	100%	13%	14%	50%		
4	15	9%		4582	6614	191	230	77	0	0%	601	16	1.01	0%	10598	36.9	100%	10.3	80%	493*	57	27%	47%	20%	40%		
5	27	16%		1997	1964	79	21	104	109	100%	4333	122	1.01	0%	16055	0.1	4%	0.8	0%	335	43	0%	59%	27%	50%		
6	11	6%		2096	2956	204	55	82	100	100%	7631	170	1.70	18%	21428	0.1	0%	1.1	9%	303	325	100%	64%	36%	45%		
7	28	16%		2355	1711	108	29	112	151	100%	3690	148	1.06	4%	17489	0.1	0%	8.2	100%	457*	55	14%	64%	37%	37%		
8	14	8%		4621	4777	487	114	168	406	100%	5588	154	1.01	0%	6790	10.7	100%	17.0	100%	349*	58	29%	93%	38%	23%		
9	11	6%		3464	4916	184	52	167	483	100%	5626	1038	6.33	100%	13131	2.0	73%	18.5	100%	633	57	18%	91%	64%	82%		
10	7	4%		4827	4806	637	59	133	257	100%	14143	988	5.69	100%	12380	0.1	0%	1.4	14%	237	43	0%	100%	71%	57%		

Legend:

	Concentration significantly higher than controls and higher than 6 or more other clusters
	Concentration significantly higher than controls and higher than 3 or more other clusters
	Concentration significantly higher than controls and higher than 2 or more other clusters
	Concentration significantly higher than controls but not higher than other clusters

FIG 3. Modified heat map of clustering of individual cases. Rows define clusters of patients with CRS; these were arbitrarily (nonhierarchically) ordered according to the cytokine pattern. Columns indicate variables used for cluster analysis, which were ordered according to their interrelationship in component analysis. Geometric mean concentrations are given for each cluster. Some variables were used in the cluster analysis in a categorical manner because of a high rate of measurements of less than the detection limit (IFN- γ , IL-5, SE-IgE, TNF- α , and IL-17). For these, proportions of positive values are provided, and additionally, geometric means are presented for illustrative purposes. Selected phenotype (proportion with nasal polyps, asthma, and allergy) is tabulated, although not used in cluster analysis. For characterization of the clusters, multiple group comparison for between-cluster differences and differences from the control group are visualized with a color code, as in the legend. For the CRSwNP proportion, only between-cluster differences were calculated. *In clusters 2, 4, 7, and 8, IL-22 concentrations were significantly higher than in clusters 1, 5, 6, and 9 but were not different from control values.

The latter 6 clusters were characterized by IL-5 expression in every single subject. These clusters also expressed increased ECP and total IgE concentrations. This IL-5-positive group could be further differentiated in a group of 3 clusters with mean concentrations of IL-5 between 100 to 151 pg/mL (moderate-expression group) and a high-expression group (257-483 pg/mL) including clusters 8 to 10. In parallel, with eosinophilic inflammation, there was neutrophilic inflammation with increased concentrations of IL-8 and MPO in clusters 6 to 10.

The IL-5 moderate-expression group (clusters 5-7) predominantly but not exclusively consisted of the CRSwNP phenotype. Further differentiation within these groups was made by coexpression of IFN- γ (cluster 6) or TNF- α and increased IL-22 levels (cluster 7). The high-expression group (clusters 8-10) was almost exclusively composed of nasal polyps and had the highest levels of albumin, IL-6, and IL-8. Additionally, clusters 8 and 10 showed a T_H17 family-based inflammation. Clusters 9 and 10 demonstrated the highest concentrations of total IgE (around 1000 kU/L) and the highest rate of comorbid asthma (ie, 64% to 71%). All samples in these 2 clusters expressed SE-IgE antibodies in tissue.

DISCUSSION

We present inflammatory endotypes of CRS based on a cluster analysis of biomarkers; secondarily, we correlate these results to clinical characteristics (phenotypes). In this cohort of European subjects who underwent surgery for CRS, we observed

considerable data variability, so that multiple homogeneous subgroups (clusters) could be differentiated. This observation indicates that CRS is not a homogeneous inflammatory disease and that endotypes are present with a wide diversity of inflammatory profiles. To our knowledge, this is the first study to report the existence of inflammatory endotypes within CRS unbiased by phenotype data. Furthermore, we demonstrated that the newly defined endotypes mirror clinical phenotypes, supporting their clinical relevance.

To determine which parameters were relevant in the generation of this diversity, principal component analysis and clustering of variables showed that diversity in our samples was driven by 5 groups of related cytokines: (1) markers of eosinophilic, T_H2-driven inflammation and antibody production (ECP, IL-5, IgE, and SE-IgE) together with albumin; (2) neutrophilic and proinflammatory cytokines (IL-1 β , IL-6, IL-8, and MPO); (3) T_H17- or T_H22-related markers (IL-17A, IL-22, and TNF- α); and (4) the T_H1 marker IFN- γ . Moreover, TGF- β 1 had a separate but small contribution to variability in our subjects.

These observations have previously been made, although in specific CRS subgroups. The correlations of ECP, IL-5, and IgE levels confirm previous data in white patients with nasal polyps, in whom inflammation is dominated by activated eosinophils¹⁵ and IgE concentrations were related to eosinophilic inflammation.²¹ The association of albumin with eosinophil products can be related to observations in nasal polyps in which increased albumin deposition has been thought to be caused by

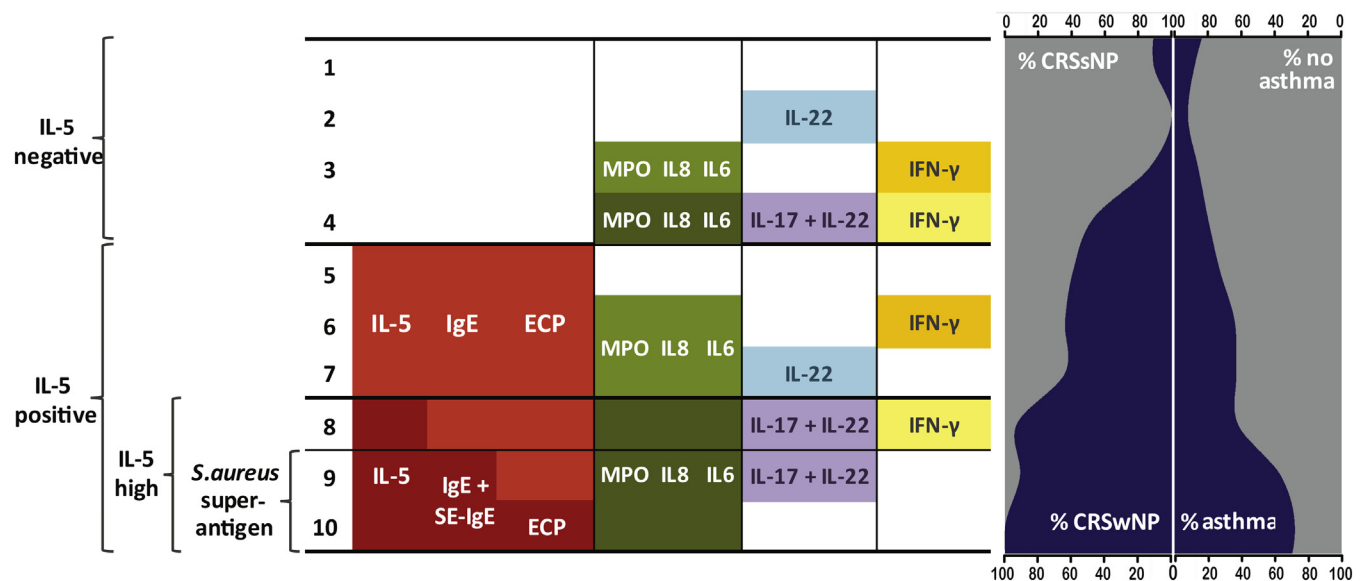


FIG 4. Summary graph. Simplified graphic depiction of the clusters and their characteristic cytokines, as well as the distribution of CRSsNP versus CRSwNP and asthma. For cytokines, white indicates no increased concentration, light colors indicate moderately increased concentrations, and dark colors indicate strongly increased concentrations. Horizontal lines indicate groups of clusters, as determined by IL-5, SE-IgE, and CRSwNP and asthma characteristics.

eosinophil-driven inflammation.¹⁵ The expression of neutrophilic and proinflammatory cytokines has been observed previously in Chinese patients with nasal polyps⁹ and more specifically in Chinese patients with nasal polyps who were IL-5/IL-17/IFN-γ negative.²²

Importantly, our observations were present in the whole CRS group, whereas previous results were limited to specific subgroups, implicating that that these previously described mechanisms can be extrapolated to the overall CRS group and might not be restricted to the CRSwNP phenotype. Indeed, separate *post hoc* analyses in the CRSsNP and CRSwNP subgroups yielded similar principal components (data not shown), indicating that even in lower concentration ranges, concentration gradients of cytokines are present and meaningful. These findings indicate that CRS is not a dichotomous or 1-dimensional linear spectrum of diseases but rather multidimensional, with different polarizations on the main inflammatory axes.

When patients were clustered into inflammatory endotypes, our analysis resulted in 10 clusters of which each cytokine pattern was interpreted, and accordingly, clusters could be grouped together (summarized in Fig 4). To characterize each cluster's phenotype, we correlated clusters with selected phenotype parameters, such as nasal polyp prevalence and asthma comorbidity. Importantly, this was a *post hoc* analysis, and the phenotype information was not used during cluster analysis.

We show a strong correlation with phenotype, with 3 clusters composed almost exclusively of patients with CRSsNP without asthma comorbidity and 3 clusters composed almost exclusively of patients with CRSwNP with strongly increased asthma prevalence. The presence of IL-5 remains the paramount factor dictating the phenotype with nasal polyps and asthma, which is in line with previous evidence.¹³ The presence of SE-IgE remains associated with nasal polyps with intense eosinophilic inflammation with very high IgE concentrations and asthma

comorbidity, as demonstrated previously.¹³ As summarized in Fig 4, clusters can be grouped into 4 IL-5–negative clusters, 3 moderately increased IL-5–positive clusters, and 3 IL-5–high clusters, of which 2 were SE-IgE–positive clusters.

However, in clusters 4 to 7 there is a mixture of phenotypes, with polyp prevalence of around 50% to 60% and moderately increased albumin levels, although each of these clusters had a distinct inflammatory profile. Three of these clusters had moderate increase in IL-5 concentrations and an eosinophilic pattern but differed in terms of IFN-γ and TNF-α expression. This could indicate that IL-5 is not the only key to edema and polyp formation for some groups and that unmeasured factors might play a role in differentiation toward a polyp. Our study was limited to the major cytokines, and the amount of tissue available precluded further analyses. Also, subjects were allocated to a phenotype during office-based endoscopy before surgery, and the clinical differentiation might be unclear in some cases in which polyps were only observed during surgery. In cases of middle turbinate–confined polypoid edema, the phenotype might not be clear cut during office-based endoscopy, as it has been shown that early-stage nasal polyps already show eosinophilic inflammation.^{15,23} Furthermore, this overlap might be inherent to the concept of phenotype, which, as has been shown in asthma phenotyping, does not provide adequate insight into the underlying pathogenic mechanisms.²⁴

Within the non- T_H2 group, inflammation is variable, from completely absent to broad and involves T_H1 , T_H17 , or T_H22 . The role for T_H1 cells in only some of the IL-5–negative subjects contrasts with earlier findings showing increased levels of IFN-γ protein and mRNA in patients with CRSsNP.⁸ However, this could be explained by the averaging effect of former analytic approaches. One of the overlapping clusters (cluster 4) had no eosinophils or IL-5 and, in contrast, had a T_H17 neutrophil-type inflammation. It is possible that this endotype is the same as what has been observed before in Chinese patients with nasal

polyps.⁹ Interestingly, it was recently shown that second-generation Asian patients with CRSwNP in the United States had a higher rate of noneosinophilic polyps than white patients.²⁵ Unfortunately, we had no ancestry information in our study to further analyze this relationship.

Production of the T_H17 family–derived cytokines IL-17A and IL-22 correlated modestly, as shown by using principal component analysis. Reasons for this modest correlation are apparent when analyzing the cluster profiles; in clusters 4 and 8 IL-17A production was paired with increased IL-22 production, whereas in clusters 2 and 7 only increased IL-22 production was present, possibly indicating T_H22 involvement in these clusters. Interestingly, in cluster 10, which was an SE-IgE–positive cluster, increased IL-22 production was paired with only marginal IL-17A production, indicating a T_H22 involvement that was also observed to be induced by staphylococcal enterotoxins in patients with atopic dermatitis.²⁶ Additionally, we observed increased TNF- α levels in all of these clusters, which is consistent with the observation that both T_H17 and T_H22 cells can induce TNF- α production.²⁷ However, IL-22 concentrations did not differ from those of control subjects, and this effect remained after excluding allergic control subjects (data not shown). The reasons for this discrepancy with the coherent between-cluster differences as previously described are not clear, and the relevance of T_H22 cells in the pathology of CRS should be further elucidated.

Surprisingly, we found little contribution of TGF- β 1 to the variability of our data; the TGF- β 1 concentration had no significant between-cluster differences and did not differ between clusters and the control group. Previously, expression of TGF- β signaling was found to be increased in white patients with CRSsNP, sharply contrasting to a decreased signaling in patients with CRSwNP.²⁸ These findings were later confirmed in Chinese patients with CRS,²⁹ although these are characterized by different T effector cell profiles than their white counterparts.⁹ In our study patients with CRSsNP were confirmed to have increased TGF- β 1 concentrations compared with patients with CRSwNP and control subjects ($P < .001$, data not shown). However, we did not have adequate amounts of tissue to further analyze remodeling patterns and regulatory T-cell involvement. Taken together, these findings might indicate that remodeling is strongly correlated with the phenotype (fibrosis vs edema formation), irrespective of the inflammatory endotype.

Some of the broad characteristics of this clustering are found in a previous cluster analysis by Nakayama et al,³⁰ who proposed a clustering in which nasal polyp scores and tissue eosinophilia mainly determined the phenotype. Interestingly, in a recent cluster analysis of patients with CRS focusing on symptoms and clinical information, rate of nasal polyposis did not differ between phenotypes.³¹ However, these approaches are inherently different from our study because we based our cluster analysis solely on inflammatory mediators to obtain endotypes, and we compared these with selected phenotype parameters, such as nasal polyp prevalence and asthma comorbidity, only in a *post hoc* analysis.

Our study has some limitations. Our definition of endotypes was entirely based on inflammatory biomarkers. However, for asthma, it is proposed that more parameters, such as clinical characteristics, genetics, and treatment responses, should be included for the definition of an endotype.³² We aimed to propose “inflammatory endotypes,” and further studies are needed that

incorporate clinical characteristics, genetics, remodeling parameters, and treatment responses to further define these endotypes. Some of these endotypes might eventually be merged or split with other endotypes given these additional studies.

The use of cluster analysis is potentially a limiting factor in this study. Because cluster analysis is not an inferential statistical technique, there is no probability calculated. The technique generates multiple possible solutions in terms of a number of clusters, from which the most relevant solution is selected by the examiner. To scrutinize this process, the selection of the best number of clusters was based on objective criteria, such as the Hubert-Levine index as a measure of internal cluster validity and bootstrap resampling of the cluster analysis. Even though all these measures resulted in a certain optimal number of clusters, it is not our aim to claim a final definite classification in which all patients with CRS should be categorized but instead illustrate the diversity in inflammatory profiles.

The inflammatory profiles of asthmatic patients have been found to evolve over time within subjects. Because we only analyzed our subjects at 1 time point, possible variability of the inflammatory endotype over time limits our findings. Indeed, when comparing the profiles of patients with nasal polyps undergoing 2 surgeries because of disease recurrence, it has been found that all nasal polyps remained IL-5 positive over an average period of 5 years; however, in some patients IL-17 was coexpressed.¹² To determine the clinical and therapeutic utility of the inflammatory endotypes, further longitudinal studies are needed, evaluating their stability over time in relation to possible treatment and environmental changes.

The applicability of our study, which was done in a European population, to other regions in the world remains at question as well. For example, the high smoking prevalence in our study was comparable with the European average, but lower smoking prevalences in North America might affect the inflammatory endotype. Other lifestyle and environmental factors might affect the endotype, and global studies are needed to address these.

Furthermore, our choice of biomarkers did not cover remodeling parameters, and we have not included cellular contents. However, the current analysis enables appreciation of the possible target groups for various therapeutic mAbs currently under study.

In summary, we have identified distinct inflammatory endotypes within patients with CRS, which largely correlated with phenotypes and further differentiated them. Endotypes clearly provide a more accurate description of the inflammatory mechanisms involved than phenotype only, and inflammation in patients with CRS might be more diverse than previously assumed. CRS-related inflammation should be considered multidimensionally heterogeneous on the T_H1, T_H2, T_H17, eosinophilic/neutrophilic, proinflammatory, superantigen, and possibly T_H22 axes. Inflammatory endotypes might be of importance for the prognosis of comorbid asthma development or disease recurrence after surgery.¹² Because recent mAb approaches, such as omalizumab,³³ mepolizumab,³⁴ or dupilumab,³⁵ target specific cytokines or pathways, it is expected that an approach based on biomarker-defined clusters would identify responders to these drugs.³⁶ For example, nasal levels of IL-5 were found to predict responders to reslizumab, an anti-IL-5 mAb.³⁷ However, the clinical applicability of inflammatory endotypes would be aided by the availability of

noninvasive serum or nasal secretion markers; a range of markers is currently under investigation in the same cohort of patients with CRS.

We thank Drs K. Van Steen and J. Mahachie for their support in statistical analysis.

Key message

- CRS consists of several inflammatory endotypes defined by the composition of inflammatory markers. Those endotypes correlate with the clinical expression of disease and asthma.

REFERENCES

- Hastan D, Fokkens W, Bachert C, Tomassen P, Newson R, Jarvis D, et al. Chronic rhinosinusitis in Europe—an underestimated disease. A GA²LEN study. *Allergy* 2011;66:1216-23.
- Pleis JR, Lucas JW, Ward BW. Summary health statistics for U.S. adults: National Health Interview Survey, 2008. *Vital Health Stat* 10 2009;242:1-157.
- Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, et al. European Position Paper on Rhinosinusitis and Nasal Polyps 2012. *Rhinology Suppl* 2012; 23:3 p preceding table of contents, 1-298.
- Jarvis D, Newson R, Lotvall J, Hastan D, Tomassen P, Keil T, et al. Asthma in adults and its association with chronic rhinosinusitis: the GA²LEN survey in Europe. *Allergy* 2012;67:91-8.
- Dietz de Loos DA, Hopkins C, Fokkens WJ. Symptoms in chronic rhinosinusitis with and without nasal polyps. *Laryngoscope* 2013;123:57-63.
- Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* 2012;18:716-25.
- Van Zele T, Claeys S, Gevaert P, Van Maele G, Holtappels G, Van Cauwenberge P, et al. Differentiation of chronic sinus diseases by measurement of inflammatory mediators. *Allergy* 2006;61:1280-9.
- Van Bruaene N, Perez-Novo CA, Basinski TM, Van Zele T, Holtappels G, De Ruyck N, et al. T-cell regulation in chronic paranasal sinus disease. *J Allergy Clin Immunol* 2008;121:1435-41, e1-3.
- Zhang N, Van Zele T, Perez-Novo C, Van Bruaene N, Holtappels G, DeRuyck N, et al. Different types of T-effector cells orchestrate mucosal inflammation in chronic sinus disease. *J Allergy Clin Immunol* 2008;122:961-8.
- Derycke L, Eyerich S, Van Crombruggen K, Perez-Novo C, Holtappels G, Deruyck N, et al. Mixed T helper cell signatures in chronic rhinosinusitis with and without polyps. *PLoS One* 2014;9:e97581.
- Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet* 2008;372:1107-19.
- Van Zele T, Holtappels G, Gevaert P, Bachert C. Differences in initial immunoprofiles between recurrent and nonrecurrent chronic rhinosinusitis with nasal polyps. *Am J Rhinol Allergy* 2014;28:192-8.
- Bachert C, Zhang N, Holtappels G, De Lobel L, van Cauwenberge P, Liu S, et al. Presence of IL-5 protein and IgE antibodies to staphylococcal enterotoxins in nasal polyps is associated with comorbid asthma. *J Allergy Clin Immunol* 2010;126: 962-8, e1-6.
- Heinzerling LM, Burbach GJ, Edenharter G, Bachert C, Bindslev-Jensen C, Bonini S, et al. GA²LEN skin test study I: GA²LEN harmonization of skin prick testing: novel sensitization patterns for inhalant allergens in Europe. *Allergy* 2009;64: 1498-506.
- Bachert C, Gevaert P, Holtappels G, Cuvelier C, van Cauwenberge P. Nasal polyposis: from cytokines to growth. *Am J Rhinol* 2000;14:279-90.
- Team RC. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2012.
- Chavent M, Simonet VK, Liqueur B, Saracco J. ClustOfVar: an R package for the clustering of variables. *J Stat Software* 2012;50:1-16.
- Baker FB, Hubert LJ. Measuring the power of hierarchical cluster analysis. *J Am Stat Assoc* 1975;70:31-8.
- Hubert LJ, Levin JR. A general statistical framework for assessing categorical clustering in free recall. *Psychol Bull* 1976;83:1072-80.
- Hennig C. Cluster-wise assessment of cluster stability. *Comput Stat Data Anal* 2007;52:258-71.
- Bachert C, Gevaert P, Holtappels G, Johansson SG, van Cauwenberge P. Total and specific IgE in nasal polyps is related to local eosinophilic inflammation. *J Allergy Clin Immunol* 2001;107:607-14.
- Ba L, Zhang N, Meng J, Zhang J, Lin P, Zhou P, et al. The association between bacterial colonization and inflammatory pattern in Chinese chronic rhinosinusitis patients with nasal polyps. *Allergy* 2011;66:1296-303.
- Meng J, Zhou P, Liu Y, Liu F, Yi X, Liu S, et al. The development of nasal polyp disease involves early nasal mucosal inflammation and remodelling. *PLoS One* 2013;8:e82373.
- Wenzel S. Severe asthma: from characteristics to phenotypes to endotypes. *Clin Exp Allergy* 2012;42:650-8.
- Mahdavinia M, Suh LA, Carter RG, Stevens WW, Norton JE, Kato A, et al. Increased noneosinophilic nasal polyps in chronic rhinosinusitis in US second-generation Asians suggest genetic regulation of eosinophilia. *J Allergy Clin Immunol* 2015;135:576-9.
- Niebuhr M, Mainard J, Heratizadeh A, Satzger I, Werfel T. Staphylococcal exotoxins induce interleukin 22 in human Th22 cells. *Int Arch Allergy Immunol* 2014;165:35-9.
- Jia L, Wu C. The biology and functions of Th22 cells. *Adv Exp Med Biol* 2014; 841:209-30.
- Van Bruaene N, Derycke L, Perez-Novo CA, Gevaert P, Holtappels G, De Ruyck N, et al. TGF-beta signaling and collagen deposition in chronic rhinosinusitis. *J Allergy Clin Immunol* 2009;124:253-9, e1-2.
- Li X, Meng J, Qiao X, Liu Y, Liu F, Zhang N, et al. Expression of TGF, matrix metalloproteinases, and tissue inhibitors in Chinese chronic rhinosinusitis. *J Allergy Clin Immunol* 2010;125:1061-8.
- Nakayama T, Asaka D, Yoshikawa M, Okushi T, Matsuaki Y, Moriyama H, et al. Identification of chronic rhinosinusitis phenotypes using cluster analysis. *Am J Rhinol Allergy* 2012;26:172-6.
- Soler ZM, Hyer JM, Ramakrishnan V, Smith TL, Mace J, Rudmik L, et al. Identification of chronic rhinosinusitis phenotypes using cluster analysis. *Int Forum Allergy Rhinol* 2015;5:399-407.
- Lotvall J, Akdis CA, Bacharier LB, Bjerrmer L, Casale TB, Custovic A, et al. Asthma endotypes: a new approach to classification of disease entities within the asthma syndrome. *J Allergy Clin Immunol* 2011;127:355-60.
- Gevaert P, Calus L, Van Zele T, Blomme K, De Ruyck N, Bauters W, et al. Omalizumab is effective in allergic and nonallergic patients with nasal polyps and asthma. *J Allergy Clin Immunol* 2013;131:110-6.e1.
- Gevaert P, Van Bruaene N, Cattaert T, Van Steen K, Van Zele T, Acke F, et al. Mepolizumab, a humanized anti-IL-5 mAb, as a treatment option for severe nasal polyposis. *J Allergy Clin Immunol* 2011;128:989-95, e1-8.
- Bachert C, Mannent L, Naclerio RM, Mullol J, Ferguson BJ, Gevaert P, et al. Effect of subcutaneous dupilumab on nasal polyp burden in patients with chronic sinusitis and nasal polyposis. A randomized clinical trial. *JAMA* 2016; 315:469-79.
- Bachert C, Zhang L, Gevaert P. Current and future treatment options for adult chronic rhinosinusitis: focus on nasal polyposis. *J Allergy Clin Immunol* 2015; 136:1431-40.
- Gevaert P, Lang-Loidolt D, Lackner A, Stammberger H, Staudinger H, Van Zele T, et al. Nasal IL-5 levels determine the response to anti-IL-5 treatment in patients with nasal polyps. *J Allergy Clin Immunol* 2006;118:1133-41.

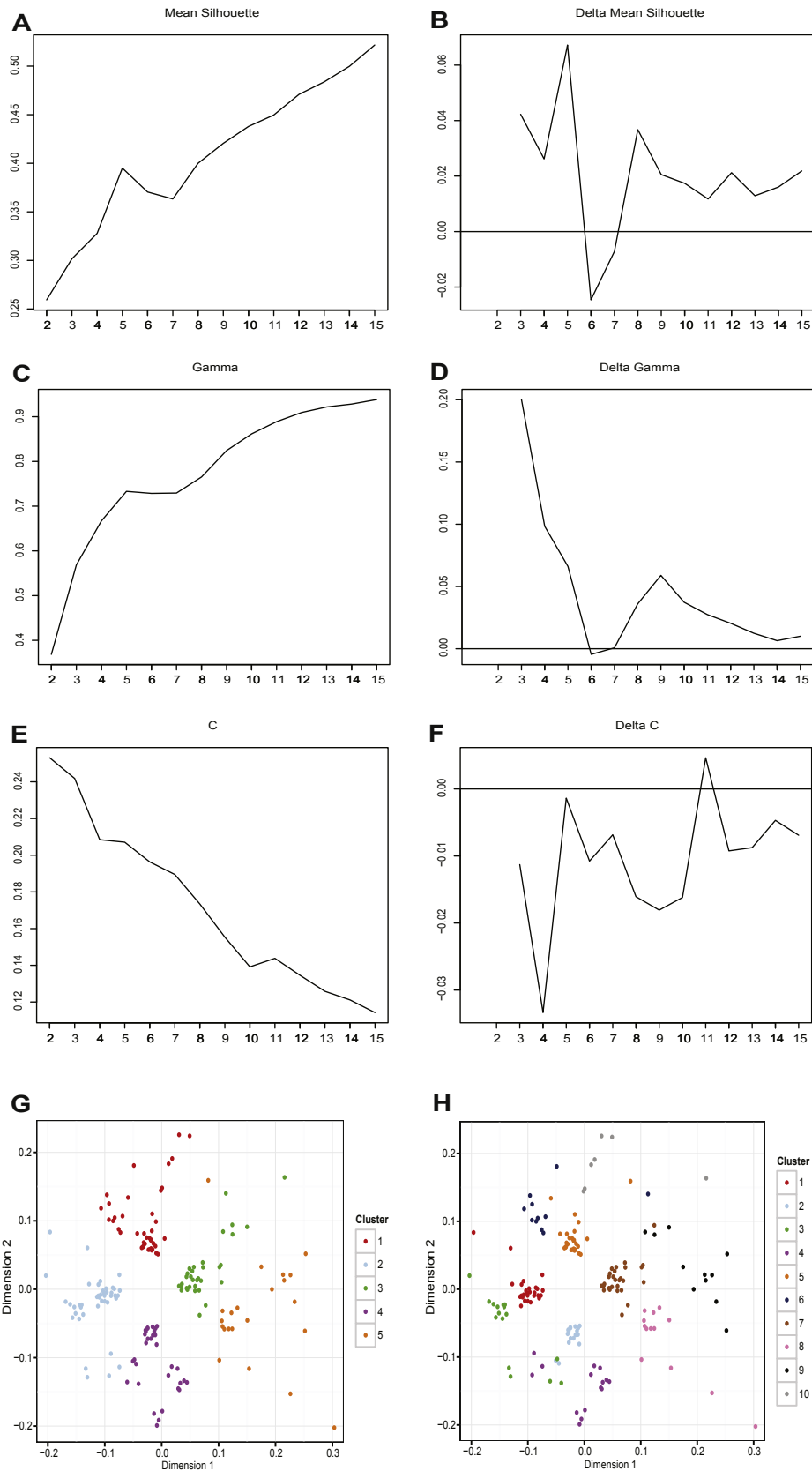


TABLE E1. Coordinates of principal component analysis for the first 5 orthogonally rotated principal components

	Component 1	Component 2	Component 3	Component 4	Component 5
IgE		0.846			
ECP		0.84			
TNF- α			0.678		
IL-8	0.876				
IL-17A	0.611		0.442		
IL-6	0.76	0.432			
IL-5		0.904			
IL-1 β	0.867				
IFN- γ					0.857
MPO	0.811				
IL-22			0.741		
TGF- β 1				0.873	
Albumin		0.69			
SE-IgE				0.423	0.47
Variance explained (%)	25	22	10	9	8
Cumulative variance explained (%)	25	47	57	66	74

Variables with coordinates of less than 0.4 were omitted from the component. The proportion of total variance in the data set, as explained by each component, is given. Also, the cumulative proportion of total variance explained by the sum of each of the components and its preceding components is given. Components indicate primary components.

FIG E1. Validation of clustering. We generated different clusters, with the number of clusters (k) ranging from 2 to 15. **A-F,** To validate clustering outcomes, we used internal cluster quality measures. For each of the possible number of clusters, an index is calculated reflecting the between-subject similarity within clusters and the dissimilarity between clusters. This index usually increases monotonically with increasing number of clusters, and the optimal value is determined to be at the elbow of its plot, where the change in index (difference with k-1 and k+1) is at a maximum. There are several possible indices available. We used mean silhouette width, the Baker-Hubert Gamma statistic, and the Hubert-Levin C index because they rely on dissimilarity data (allowing mixed continuous and categorical data, as is the case in our sample). We mostly relied on the C index because this is especially fit for mixed-type data. The following plots show the index for k of 2 to 15 possible clusters and the change (delta) in index compared with k-1. Note that for the C index, lower values are better. The results from the internal cluster quality indexes produce a good signal at either 4 to 5 or 8 to 10 clusters. Second, we assessed clustering stability after resampling of the cluster analysis. Here, data are resampled (for 1000 iterations) by using several schemes (bootstrap and subsetting of the data) and clusters are recalculated. The Jaccard similarities of the original clusters to the most similar clusters in the resampled data are computed, producing an estimate of the stability of a cluster. In the range of k = 2 to 15, average stability peaked at 5 and 10 clusters. Lastly, we assessed cluster validity using visual inspection of the clusterplot. The subjects are plotted in a 2-dimensional space after multidimensional scaling, which is a technique maximizing the dissimilarities between subjects projected in 2 dimensions. **G** and **H,** Plots were created for the 5- and 10-cluster solution. The 5-cluster solution clearly showed unresolved clusters, which were correctly discovered in the 10-cluster solution.

TABLE E2. Bootstrap stability of individual clusters in the 10-cluster solution

Cluster	Bootstrap stability
1	0.98
2	0.95
3	0.82
4	0.87
5	0.96
6	0.70
7	0.94
8	0.77
9	0.61
10	0.68

Data were resampled (for 1000 iterations) by using several schemes (bootstrap and subsetting of the data), and clusters were recalculated. The Jaccard similarities of the original clusters to the most similar clusters in the resampled data are computed, producing an estimate of the stability of a cluster. In general, values of less than 0.5 indicate unstable clusters, values of greater than 0.6 indicate plausible structure, values of greater than 0.75 indicate stable and valid clusters, and values of greater than 0.85 indicate highly stable clusters.

TABLE E3. Clinical characteristics of clusters

Cluster	Ever smoker (%)	Current smoker (%)	Aspirin sensitivity (%)	Sex (% female)	Previous surgery rate (%)
1	72	40	0	28	0
2	81	56	0	31	21
3	73	53	0	27	8
4	87	33	7	53	27
5	46	8	11	35	39
6	73	18	9	73	30
7	69	31	7	54	33
8	64	14	7	29	39
9	73	36	18	36	50
10	14	0	0	14	64