

DR MICHAEL FRICKER (Orcid ID : 0000-0002-8587-1774)

Article type : Original Article-Asthma and Rhinitis

Title: Dysregulation of sputum columnar epithelial cells and products in distinct asthma phenotypes

Authors: Ling Qin, MD^{1,2}, Peter G. Gibson, MBBS FRACP^{1,3,4,5}, Jodie L. Simpson, PhD^{1,3,4}, Katherine J. Baines, PhD^{1,3,4}, Vanessa M. McDonald, PhD^{1,3,4,5}, Lisa G. Wood, PhD^{1,3,4}, Heather Powell, MMedSci^{1,4}, Michael Fricker, PhD^{1,3,4}

Running title: Sputum measures of epithelial dysregulation in asthma

Author affiliations: ¹Priority Research Centre for Healthy Lungs, The University of Newcastle, NSW, Australia.

²Department of Respiratory Medicine (Department of Pulmonary and Critical Care Medicine), Xiangya Hospital, Central South University, China.

³National Health and Medical Research Council Centre of Excellence in Severe Asthma, Newcastle, NSW, Australia.

⁴Hunter Medical Research Institute, Newcastle, NSW, Australia.

⁵Department of Respiratory and Sleep Medicine, John Hunter Hospital, Newcastle, NSW, Australia.

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.13452

This article is protected by copyright. All rights reserved.

Name and address of corresponding author: Michael Fricker

Address: Priority Research Centre for Healthy Lungs, Hunter Medical Research Institute, Lot 1 Kookaburra Circuit, New Lambton Heights, NSW, 2305, Australia.

Phone: +61 2 404 20207

Fax: +61 2 404 20046

Email: michael.fricker@newcastle.edu.au

Author contributions: L.Q. designed and executed studies, performed analysis, interpreted data and wrote the manuscript. P.G.G. designed and supervised studies, contributed retrospective data, interpreted data and edited the manuscript. J.L.S., K.J.B., V.M.M. and L.G.W. contributed retrospective cohort data, interpreted data and edited the manuscript. H.P. performed statistical analysis and edited the manuscript. M.F. designed, executed and supervised studies, performed analysis, interpreted data and wrote the manuscript.

Key words: Severe asthma, sputum, epithelium, biomarker, inflammation, transcriptomic, eosinophil, epithelial cell, periostin, gene signature

Abbreviations:

ACQ	Asthma Control Questionnaire
AHR	Airways Hyperresponsiveness
AUC	Area Under the Curve
BMI	Body Mass Index
CEP	Columnar Epithelial Cell
CFAP	Cilia and Flagella Associated Protein
CLCA1	Calcium-Activated Chloride Channel Protein 1
DNALI1	Dynein Axonemal Light Intermediate Chain 1

EA	Eosinophilic Asthma
EPCAM	Epithelial Cellular Adhesion Molecule
FENO	Fractional Exhaled Nitric Oxide
FEV₁	Forced Expiratory Volume in 1 second
FOXJ1	Forkhead box protein J1
FVC	Forced Vital Capacity
GO	Gene Ontology
HC	Healthy Control
ICS	Inhaled Corticosteroid
IL	Interleukin
LDLRAD1	Low Density Lipoprotein Receptor Class A Domain Containing 1
PD15	Dose of saline required to provoke 15% fall in FEV ₁
PGA	Paucigranulocytic Asthma
ROC	Receiver Operating Characteristic
SCGB3A1	Secretoglobin Family 3A Member 1
SD	Standard Deviation
TTC25	Tetratricopeptide Repeat Domain 25

Abstract

Background: Dysfunction of the bronchial epithelium plays an important role in asthma, however its measurement is challenging. Columnar epithelial cells are often quantified, yet rarely analysed, in induced sputum studies.

Objective: We aimed to test if sputum columnar epithelial cell proportion and count are altered in asthma, and if they are associated with clinical and inflammatory variables. We aimed to test whether sputum-based measures could provide a relatively non-invasive means through which to monitor airway epithelial activation status.

Methods: We examined the relationship of sputum columnar epithelial cells with clinical and inflammatory variables of asthma in a large retrospective cross-sectional cohort (901 participants with asthma and 138 healthy controls). In further studies we used flow cytometry, microarray, qPCR and ELISA to characterise sputum columnar epithelial cells and their products.

Results: Multivariate analysis and generation of 90th centile cut-offs ($\geq 11\%$ or $\geq 18.1 \times 10^4/\text{mL}$) to identify columnar epithelial cell “high” asthma revealed a significant relationship between elevated sputum columnar cells and male gender, severe asthma and non-neutrophilic airway inflammation. Flow cytometry showed viable columnar epithelial cells were present in all sputum samples tested. An epithelial gene signature (*SCGB3A1*, *LDLRAD1*, *FOXJ1*, *DNALI1*, *CFAP157*, *CFAP53*) was detected in columnar epithelial cell-high sputum. *CLCA1* mRNA and periostin protein, previously identified biomarkers of IL-13-mediated epithelial activation, were elevated in columnar epithelial cell-high sputum samples, but only when accompanied by eosinophilia.

Conclusions & Clinical Relevance: Sputum columnar epithelial cells are related to important clinical and inflammatory variables in asthma. Measurement of epithelial biomarkers in sputum samples could allow non-invasive assessment of altered bronchial epithelium status in asthma.

Introduction

Induced sputum analysis is a widely published analytical technique in clinical studies of chronic inflammatory airways diseases. Although relatively technically challenging, the principal advantage of sputum analysis over biopsy and bronchoalveolar lavage is that it is the least invasive method available to directly sample and measure inflammatory changes in the airways.

Conventional sputum analysis involves microscope-based counting of immune cells. Sputum eosinophilia has received particular attention, as a biomarker of type 2 inflammatory pathways and corticosteroid response in asthma, and due to its association with higher risk of severe exacerbations[1-4]. Transcriptomic and proteomic analyses of sputum have further refined our understanding of asthma airway inflammatory phenotypes[5-8]. We have demonstrated that sputum gene expression signatures can predict airway inflammatory phenotype and inhaled and oral corticosteroid response[9, 10].

Columnar epithelial cells (CEPs) are present in sputum samples and are routinely quantified. The bronchial epithelium is a critical site of asthma-related pathophysiological changes, and transcriptomic studies of bronchial brushings demonstrate altered bronchial epithelium activation status in asthma[11-13]. A three gene expression signature (composed of periostin, CLCA1 and SerpinB2) measured in bronchial brushings relating to IL-13 activation of

bronchial epithelial cells, type 2 high asthma and airway eosinophilia can predict corticosteroid response in patients with mild asthma[13, 14]. The use of epithelial gene signatures in the clinic has been limited by the highly invasive nature of bronchial brushing collection. The epithelial gene signature components periostin, CLCA1 and SerpinB2 have been detected in sputum samples, where their significance is poorly understood[15]. Periostin protein, released at the basolateral surface of the bronchial epithelium[16], has shown promise as a serum-based biomarker of type 2 inflammation in clinical trials of IL-4/13-targeted biologicals[17]. Despite its basolateral-directed secretion, periostin protein is also detectable in sputum supernatants[18], and both serum and sputum periostin measurements provide significant but modest prediction of airway eosinophilia[18, 19].

Despite their measurement, the potential significance of sputum CEP measures from a clinical and biomarker perspective has not been investigated. We hypothesized that sputum CEP proportion and count would be altered in asthma, and would be associated with clinical and inflammatory variables. We further hypothesized that sputum CEPs could provide a relatively non-invasive means through which to monitor airway epithelial activation status.

Methods

Study populations

We performed a retrospective cross-sectional study on 901 participants with asthma and 138 healthy controls (HC) (cohort 1), recruited for clinical studies at the Hunter Medical Research Institute and John Hunter Hospital, Newcastle, Australia between 2005-2015[10, 20-26]. The sole inclusion requirement for our retrospective analysis was availability of a sputum differential cell count. Cohort 1 included 83 participants with asthma in whom serum and/or sputum periostin protein was measured[18] and 59 participants with asthma for whom

sputum microarray transcriptomic data was available (Figure 1)[5]. qPCR-based validation of microarray data and measurement of periostin, SerpinB2 and CLCA1 gene expression was performed in a separate cohort (cohort 2) of 60 participants with asthma (Supplemental table E1 provides clinical and inflammatory characteristics of cohort 2 and 3). Sputum flow cytometry was performed on samples collected from an additional 17 participants with asthma (cohort 3) (Supplemental table E1).

In cohorts 1, 2 and 3 asthma was diagnosed according to ATS guidelines: current (past 12 months) episodic respiratory symptoms, doctor's diagnosis (ever), evidence of airway hyperresponsiveness (AHR) to hypertonic saline (high dose saline PD15 <15 mg/mL) or reversible airflow limitation (postbronchodilator increase in FEV₁ of $\geq 12\%$ and >200mL).

Study participants had not had a recent (past month) respiratory tract infection or asthma exacerbation. In all studies data were collected using standardized procedures and included medical history and demographics (age, gender, height, weight, smoking status and pack/years), Juniper Asthma Control questionnaire (ACQ), spirometry (pre-bronchodilator FEV₁ and FEV₁/FVC), saline challenge, allergy skin prick test or serum atopic test and fractional exhaled nitric oxide (FENO) measurement. For detailed definitions of clinical variables measured see online repository. Healthy controls had no lifetime history of pulmonary disease and lacked AHR. Study participants provided written informed consent. All studies and procedures were reviewed and approved by the Hunter New England Health human research ethics committee (Hunter New England Health ethics protocols 08/11/19/3.03, 16/04/20/3.01, 06/10/25/5.03, 05/03/09/3.09, 09/05/20/5.07, 09/03/18/5.05, 11/06/15/3.03, 07/10/24/5.03, 03/09/10/3.12, 02/02/13/3.10, 05/12/07/3.11, 08/08/20/3.10) and conformed to the standards of the Declaration of Helsinki. When appropriate, studies were registered with the Australian New Zealand Clinical Trials Registry

(ACTRN12609000197235, ACTRN12614001003662, ACTRN12607000236493, ACTRN12606000286549, ACTRN12611000235909, ACTRN12605000318684, ACTRN12611001278921, ACTRN12612000807853). The data are not publicly available due to privacy, ethical or third party restrictions.

Statistical analysis

Statistical analyses were performed with SPSS 24 software and GraphPad 7. The Skewness and Kurtosis tests were used to test the normality of distribution. Continuous variables are presented as means (SD) (parametric data) or median (interquartile range) (non-parametric data). Categorical variables are presented as frequencies and percentages. Two-group comparisons were made with independent T-test (parametric data) or Mann-Whitney test (non-parametric data) or Chi-square test (categorical data). Association of CEPs with clinical and inflammatory variables was performed by univariate and multivariate linear regression (continuous variables). For linear regression non-parametric variables were normalized using rank cases. Variables with p values <0.2 in univariate analysis were included in multivariate analysis, along with age. Factors related to the dependent variable (e.g. sputum cell proportions for CEP %) or multiple related variables were excluded (e.g. ICS dose and asthma control related to severe asthma). P values <0.05 were taken as statistically significant. For gene signature biomarker studies, predicted values were generated in multivariate logistic regression for use in ROC analysis. Multiple logistic regression with conditional backwards stepwise analysis was conducted on the sputum epithelial gene signature until all transcripts remaining had p<0.05 in relationship to CEP%. Bland-Altman analysis was performed to verify level of agreement between flow-cytometry- and microscope-based sputum CEP quantification.

Additional methods

See the online supplemental information for detailed description of sputum processing, sputum flow cytometry, periostin ELISA, sputum RNA extraction, microarray and qPCR analysis.

Results

Sputum CEPs are increased in asthma.

In cohort 1 we compared demographic, clinical, spirometry and sputum cell count data from 901 adult participants with asthma and 138 HCs. Participants with asthma were significantly older, had higher mean BMI, a higher rate of atopy, worse lung function and less were current or former smokers (Table 1). Total sputum cell counts, sputum neutrophil and eosinophil count and proportion were increased compared to HCs. The absolute number of sputum CEPs was increased in asthma (HC 3.31×10^4 /mL median, 0.96-9.95 q1,q3; Asthma 5.47×10^4 /mL median, 1.38-18.10 q1,q3; $p=0.002$), and the CEP proportion showed a similar trend (HC 1.25% median, 0.25-3.06 q1,q3; Asthma 1.75% median, 0.5-4.75 q1,q3; $p=0.057$).

Confirmation of sputum CEP identity and viability.

In our retrospective database (cohort 1), sputum CEPs were quantified in microscopy studies based on elongated morphology with extrusion of a cellular ‘tail’ on one side of the nucleus (Figure 2A). To confirm the identity of microscope-quantified CEPs, in cohort 3 we employed flow cytometry on sputum cell suspensions from participants with asthma, with CD45 negative, EpCAM positive cells ascribed as CEPs (Supplemental Fig E1 depicts the flow cytometry gating strategy used)[27]. The proportion of sputum CEPs in flow cytometry was expressed relative to total of CEPs and CD45 positive events (leukocytes) combined, analogous to the microscopy method. A strong correlation of CEP proportion measured with

the two techniques was observed ($r=0.85$, $p<0.001$), and Bland-Altman analysis showed good agreement between flow cytometry and microscopy-based CEP quantification (Figure 2B-C). Thus, flow cytometry confirmed the verity of our prior microscope-based quantification of CEPs. In a subset of flow cytometry sputum samples from cohort 3 we included viability dye 7-AAD (Supplemental Fig E1). CEP viability varied between samples, although a proportion of CEPs retained plasma membrane integrity in all samples (Figure 2D). There was no significant correlation between the percentage ($r=0.01$, $p=0.98$) or absolute number ($r=0.16$, $p=0.7$) of CEPs and their viability (Figure 2E-F).

Association of sputum CEPs with demographic, clinical and inflammatory parameters in asthma.

In cohort 1 we performed univariate analysis to examine the relationship between sputum CEP percentage (Table 2) or absolute number (Supplemental table E2 – univariate and multivariate regression analysis for sputum CEP count) and demographic, clinical and inflammatory variables. Both CEP percentage and number were significantly associated with male gender, poorly controlled asthma, severe asthma and negatively correlated with neutrophilic airway inflammation. Multivariate analysis of CEP percentage with related factors (those with $p<0.2$ in the univariate analysis, plus age) confirmed significant relation with male gender, never smoking, severe asthma, and an inverse relationship with sputum neutrophil count. Similar results were observed in multivariate analysis with sputum CEP number (Supplemental table E2).

Establishment of a sputum CEP cut-off and comparison of CEP high vs low asthma.

We established a cut-off value for normal range sputum CEP proportion (CEP % high = $\geq 11\%$) or number (CEP count high = $\geq 18.1 \times 10^4/\text{mL}$) using the 90th percentile of the cohort 1 HC population (Table 3). CEP high asthma had increased baseline FEV₁/FVC (CEP % low 70.27 vs high 74.13, $p=0.013$), an increased proportion of participants on high dose ICS (CEP % low 40.95% vs high 52.08%, $p=0.037$), and an increased proportion with severe asthma compared to the CEP low group (CEP % low 38.39% vs high 51.04%, $p=0.017$) (Table 3). Total sputum cell count (CEP % low 3.83×10^6 vs high 2.94×10^6 , $p=0.004$), lymphocyte number (CEP % low 2.91×10^4 vs high 1.91×10^4 , $p=0.018$), neutrophil number (CEP % low 138.72×10^4 vs high 48.0×10^4 , $p<0.001$), eosinophil number (CEP % low 3.83×10^4 vs high 2.94×10^4 , $p=0.01$) and neutrophil percentage (CEP % low 37.75% vs high 17.31%, $p<0.001$) were lower in high CEP participants, a higher proportion of CEP % high were paucigranulocytic compared to CEP % low participants (CEP % low 41.24% vs high 57.29%, $p=0.003$) and a lower proportion of CEP % high were neutrophilic (CEP % low 16.27% vs high 2.08%, $p<0.001$). Age, BMI and atopy were not different between the two groups, supporting the prior regression analysis. Similar results were observed when participants were categorized as high CEP or low CEP based on absolute count (Supplemental table E3 provides clinical and inflammatory characteristics of CEP count high asthma in cohort 1).

Epithelial gene signatures are elevated in sputum of CEP high asthma.

We performed an exploratory analysis of published sputum microarray data available in 59 participants with asthma from cohort 1[5], identifying differentially expressed genes (>1.5 -fold change) between CEP high ($n=5$) and low ($n=54$) asthma sputum samples (Supplemental table E4 provides sputum cell counts for CEP high and low samples analysed by microarray).

In total, 63 differentially expressed genes (adjusted $P < 0.05$) were identified, of which 62 were increased in CEP high samples (Supplemental table E5 – list of differentially expressed genes with normalized microarray values). Seven gene ontology (GO) categories were significantly overrepresented amongst the differentially expressed genes, including cilium movement and microtubule activity, supporting the relevance of the differentially expressed genes to CEP function (Supplemental table E6 – list of GO categories overrepresented in CEP high samples). We selected 7 transcripts (*SCGB3A1*, *LDLRAD1*, *FOXJ1*, *DNALI1*, *TTC25*, *CFAP53* and *CFAP157*) for further analysis by qPCR in a separate validation cohort, of which 29 samples were derived from CEP high and 31 samples from CEP low participants with asthma (cohort 2) (Supplemental table E7 – details of transcripts selected for qPCR validation study, Supplemental table E8 – sputum cell counts for CEP high and low patients in cohort 2). All 7 genes had been previously shown by bioinformatic analysis to be preferentially expressed in specific epithelial cell types. *TTC25* was undetectable by qPCR. *SCGB3A1*, *LDLRAD1*, *FOXJ1*, *DNALI1*, *CFAP53* and *CFAP157* expression were significantly elevated in CEP high samples (Fig 3A-F, Supplemental table E9 – Average cycle threshold (Ct) values for target genes in qPCR study). Multiple logistic regression (Supplemental table E10 – regression and ROC analysis of sputum qPCR and periostin protein) and ROC analysis showed the 6 gene combination significantly predicted sputum CEP category (AUC 84.4%, $p < 0.0001$) (Fig 3G). Multiple logistic regression with stepwise backwards selection identified *DNALI1* as the best individual predictor of sputum CEP status (AUC 83.1%, $p < 0.0001$) (Fig 3G and Supplemental table E10).

Epithelially-derived type 2 asthma biomarkers are selectively increased CEP high, eosinophil high sputum samples.

We examined sputum gene expression of *Periostin*, *CLCA1* and *SerpinB2* in our qPCR validation cohort (cohort 2). *Periostin* gene expression was undetectable. Neither *CLCA1* nor *SerpinB2* were increased in CEP high samples compared to CEP low (Fig 4A-B). We further categorized samples into four groups based on the presence or absence of high or low levels of sputum CEPs (CEP high, $\geq 11\%$) and/or eosinophils (eosinophil high, $\geq 3\%$). *CLCA1* expression was significantly higher in CEP high, eosinophil high samples compared to all other groups, whilst *SerpinB2* expression did not differ (Fig 4C-D). To examine the relationship of periostin with sputum CEPs and eosinophilia, we re-analysed our previously published sputum supernatant and serum periostin ELISA data[18]. Neither sputum nor serum periostin protein differed between CEP high and low samples (Fig 4E-F). Sputum (but not serum) periostin was significantly increased in CEP high, eosinophil high samples compared to all other groups (Fig 4G-H). In ROC analysis, sputum periostin protein did not significantly predict CEP high vs CEP low samples (AUC = 63.1%, $P = 0.09$) (Fig 4I). Of note, sputum periostin protein significantly distinguished high CEP high eosinophil samples from the three other groups (AUC = 83.7%, $P = 0.0009$) (Fig 4J), with the highest biomarker performance attained for distinction of CEP high eosinophil high samples from CEP low eosinophil low samples (AUC = 88.0%, $P = 0.0001$) (Supplemental Fig E2A – ROC curves for serum and sputum periostin protein and sputum *CLCA1* mRNA). In contrast, serum periostin measurement was ineffective in all equivalent analyses performed (Supplemental Fig E2B-D). Sputum *CLCA1* gene expression provided significant capacity to distinguish CEP high eosinophil high individuals from all other groups (AUC = 74.2%, $P = 0.01$) and from CEP low eosinophil low individuals (AUC = 81.4%, $P = 0.007$) (Supplemental Fig E2E-F).

Discussion

In this study, we show for the first time that variation in sputum CEP proportion and number relates to important clinical and inflammatory aspects of asthma. Using a large database of extensively characterized participants with asthma and healthy controls, we established cut-off values for normal range sputum CEP proportion and number. Patients with asthma characterized by abnormally high sputum CEP proportion or number were more likely to have severe asthma, be male, and displayed predominantly paucigranulocytic or eosinophilic airway inflammatory phenotypes. We used flow cytometry of sputum to confirm the accuracy of conventional microscope-based sputum CEP quantification, and furthermore demonstrated that sputum CEP viability is variable, but that all sputum samples contain at least a portion of viable CEP. In two separate populations we first identified and then validated a unique transcriptional signature elevated in CEP high sputum samples which had good biomarker potential. Finally, epithelial gene signatures or epithelium-derived biomarkers reflecting type 2 inflammatory epithelial responses showed potential as biomarkers in sputum, highlighting the potential of sputum analysis as a relatively non-invasive means of assessing bronchial epithelial dysfunction in chronic inflammatory airways diseases such as asthma.

The observation of increased bronchial epithelial cells in the sputum of patients with asthma dates back at least a century, with the description of large compact clusters of columnar epithelial cells or “Creola bodies”[28, 29]. Despite this, very few studies (if any) have focused on the relationship of sputum CEPs with clinical and inflammatory characteristics of asthma, and thus we sought to address this knowledge gap. Our present findings demonstrated for the first time an association of elevated sputum CEPs with the paucigranulocytic airway inflammatory phenotype in asthma and a negative association with neutrophilic asthma, and revealed a relationship between elevated sputum CEPs with poor

asthma control, high dose ICS and consequently, severe asthma. A similar retrospective cross-sectional study of 833 patients with asthma reported that sputum CEP number was increased in eosinophilic and paucigranulocytic (i.e. non-neutrophilic) patients compared to healthy controls, a finding which tallies with the negative relationship between sputum CEPs and neutrophilic airway inflammation reported here[30].

What is the explanation for increased presence of CEPs in the airway lumen of patients with asthma? The relationship between increasing sputum CEP proportion and number with poor asthma control, increased ICS dose and severe asthma suggest that this metric reflects some aspect of airway epithelial perturbation that becomes more common with increasing asthma severity. We cannot exclude that CEPs appear in induced sputum samples as a result of the sample collection process, although they are also present in spontaneous sputum samples. Regardless, the data indicate an underlying alteration in patients with asthma that contribute to the increased appearance of CEPs in sputum samples. Controversy exists regarding epithelial desquamation detected in bronchial biopsies in asthma, and whether it represents a true pathological alteration or an artefact of the sampling process[31]. However, detailed histological, cell culture and ‘omic’ studies of the bronchial epithelium in asthma, particularly severe asthma, support the notion that cellular attachment processes are disrupted, contributing to increased epithelial cell detachment and the induction of a state of aberrant epithelial layer growth and repair[32, 33].

Another possibility is the increased CEPs in the airway lumen in asthma may reflect increased apoptosis or detachment-induced death termed anoikis[34]. Consistent with this notion, and our observation that increased sputum CEP proportion and count were related to high ICS dose, are observations that CS can induce apoptosis of bronchial epithelial cells[35].

Our flow cytometry-based CEP viability measurements, although not able to detect early stages of apoptosis, would suggest that cell death was not a consistent causative factor in driving increased sputum CEPs, as there was no correlation between CEP viability and sputum CEP number or proportion. In microscopy-based differential cell counts, sputum CEPs were identified and counted only when they presented with normal cellular and nuclear morphology, suggesting that cell death may not have driven their detachment. We also discovered a unique sputum transcriptomic gene signature associated with CEP high status, including many transcripts encoding molecules with functional relevance to epithelial cells, in particular cilia function. FOXJ1 (Forkhead box J1) encodes a transcription factor that plays a major role in transcriptional regulation of motile ciliogenesis[36]. Its downstream transcriptional targets include DNALI1 (dynein axonemal light intermediate chain 1), a component of outer dynein arm, and CFAP157 and CFAP53 (cilia and flagella-associated proteins -157 and -53). TTC25 encodes a component of the outer dynein arm docking complex of cilia. SCGB3A1 (encoding secretoglobin 3A member 1) is transcriptionally enriched in the small airway epithelium and relates to clara cell function[37]. LDLRAD1 has been identified as a ciliary gene[38]. We found that DNALI1 mRNA, encoding an axonemal dynein component of cilia, significantly predicted CEP high status. Given the upregulation of epithelial-related transcripts in CEP high samples it is unlikely that epithelial cell detachment was predominantly driven by apoptosis, as induction of apoptosis is associated with rapid and extensive mRNA decay[39].

Transcriptomic analysis of bronchial brushings and biopsies indicates that considerable bronchial epithelial signaling heterogeneity exists in asthma and severe asthma[11, 13, 33, 40]. A major challenge ahead is the development of clinically-applicable biomarkers that reflect these varied alterations of epithelial function. One example of such a biomarker is

serum periostin, which has shown some potential in treatment selection/response for IL-4/13 targeted biological therapies in severe asthma[17]. Our present findings provide a plausible explanation for the detection of basolaterally-secreted periostin in the airway lumen[18]. We propose that periostin detected in sputum reflects its secretion into the airway lumen by IL-13-activated, detached sputum CEPs. The lower levels of periostin in sputum (approximately 50-fold lower) compared to serum likely reflects the limited numbers of sputum CEPs compared to the contribution of the entire bronchial epithelium (and other extrapulmonary systems) to serum periostin concentration. We also propose that the fact that elevated sputum periostin was only observed in patients with both high sputum CEP and high eosinophils indicates that it reflects the immune signaling activity of luminal CEPs. Our finding that gene expression of *CLCA1* is increased only in CEP high eosinophil high samples supports the conclusion that sputum CEPs retain the characteristic immune signaling alterations of attached bronchial epithelial cells. This along with our discovery of a unique sputum transcriptomic signature reflecting CEP high status further supports the development of sputum gene signatures as phenotyping tools in chronic inflammatory airways disease[9, 15, 41]. Thus, monitoring of epithelial gene signatures or related products in sputum samples could provide a means to monitor bronchial epithelial immune processes and potentially integrity in asthma without the requirement of bronchoscopy-mediated epithelial sampling. The obvious caveat to this statement is that these measures would only likely be informative in the subset of asthma patients in which CEPs are elevated in sputum.

This study has strengths and limitations. The retrospective nature of our initial cohort study, which included clinical details from several separate research studies conducted at our site, may have introduced a bias towards particular subtypes of asthma, for example severe asthma (40% of our cohort). Conversely, the large size of our retrospective cohort adds strength to

our findings and it is notable that similar initial findings regarding sputum CEPs and association with PGA and EA inflammatory phenotypes were previously reported[30]. Our study of sputum periostin levels was a secondary analysis of published data[18], and further validation of sputum periostin as a biomarker of CEP high, eosinophil high status would require prospective recruitment of a patient cohort. Our microarray analysis was performed using existing data and included a limited number of samples from participants with CEP high asthma, which may have limited our power to discover CEP high signature genes. Despite this, our qPCR validation study confirmed the results of this microarray analysis. Further studies using larger numbers of transcriptomic profiles of CEP high samples, and analysis of epithelial signatures in matched sputum and bronchial brushing samples, are warranted to further characterize potential epithelial alterations present and extend our observations. We observed using flow cytometry in a limited number of samples that viability of sputum CEPs was varied, and a larger study testing whether CEP viability impacts on relation to clinical and inflammatory outcomes in asthma would be of interest. Based on our cohort data we cannot exclude the possibility that smoking effects sputum CEPs and it will be important to test this in future studies.

In summary, we describe for the first time a detailed analysis of the association between sputum CEP proportion and number with demographic, clinical and inflammatory variables in asthma, revealing novel associations with factors relating to asthma severity and airway inflammation. We demonstrate that sputum CEPs appear to be biologically active and may retain and reflect the immune activation status of the bronchial epithelium, which may in turn have important mechanistic and biomarker potential in asthma.

Acknowledgements: This study was funded by the National Health and Medical Research Council of Australia (project identifier 1078579, and prior studies were funded through project identifiers 569246, 510727, APP1081533). Periostin ELISA assays were funded by Genentech. L.Q. was supported by a National Key Clinical Specialty grant from the Department of Respiratory Medicine, Xiangya Hospital, Central South University. M.F. holds a National Health and Medical Research Council of Australia Centre for Research Excellence in Severe Asthma research fellowship and a Thoracic Society of Australia and New Zealand/AstraZeneca Respiratory Disease research fellowship. We would like to thank clinical and processing team members at HMRI who performed patient recruitment and sputum analysis: Kellie Fakes, Michelle Gleeson, Naomi Fibbens, Lakshitha Gunawardhana, Bridgette Donati, Noreen Bell, Joanne Howe, Catherine Delahunty, Erin Harvey, Calida Garside, Kelly Steel, Gabrielle LeBrocq, Amber Smith, Michelle Rostas, Laura Cordova Rivera, Joanne Smart, Bronwyn Berthon and Hayley Scott. We would also like to thank the AMAZES clinical trial site leaders (Ian Yang, John Upham, Paul Reynolds, Sandra Hodge, Alan James, Christine Jenkins, Matthew Peters, Guy Marks, Melissa Baraket) and processing teams.

Conflict of interest statement: L.Q. reports grants from NHMRC and Xiangya Hospital during the conduct of the study and declares no conflict of interest. P.G.G. has received research and fellowship funding from the NHMRC, research funding from AstraZeneca, GlaxoSmithKline, and Novartis, and speaker fees from AstraZeneca, GlaxoSmithKline, and Novartis, unrelated to the current manuscript. J.L.S. declares no conflict of interest in relation to this paper. K.J.B. has received research funding from the NHMRC and fellowship funding from Lung Foundation Australia and declares no conflict of interest in relation to this paper. V.M.M. reports grants from Hunter Medical Research Institute, John Hunter Hospital

Charitable Trust, University of Newcastle during conduct of the study as well as grants and personal fees from GlaxoSmithKline and AstraZeneca and personal fees from Menarini outside of the submitted work, and declares no conflict of interest. L.G.W and H.P. declare no conflict of interest in relation to this paper. M.F. has received research and fellowship funding from the NHMRC, Thoracic Society of Australia and New Zealand and AstraZeneca, and declares no conflict of interest in relation to this paper.

References

1. Green RH, Brightling CE, McKenna S, Hargadon B, Parker D, Bradding P, et al. Asthma exacerbations and sputum eosinophil counts: a randomised controlled trial. *Lancet*. 2002;360(9347):1715-21.
2. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med*. 2008;178(3):218-24.
3. Denlinger LC, Phillips BR, Ramratnam S, Ross K, Bhakta NR, Cardet JC, et al. Inflammatory and Comorbid Features of Patients with Severe Asthma and Frequent Exacerbations. *Am J Respir Crit Care Med*. 2017;195(3):302-13.
4. Jayaram L, Pizzichini MM, Cook RJ, Boulet LP, Lemiere C, Pizzichini E, et al. Determining asthma treatment by monitoring sputum cell counts: effect on exacerbations. *Eur Respir J*. 2006;27(3):483-94.
5. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. *J Allergy Clin Immunol*. 2011;127(1):153-60, 60 e1-9.
6. Yan X, Chu JH, Gomez J, Koenigs M, Holm C, He X, et al. Noninvasive analysis of the sputum transcriptome discriminates clinical phenotypes of asthma. *Am J Respir Crit Care Med*. 2015;191(10):1116-25.

7. Kuo CS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, et al. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur Respir J*. 2017;49(2).
8. Lefaudeux D, De Meulder B, Loza MJ, Peffer N, Rowe A, Baribaud F, et al. U-BIOPRED clinical adult asthma clusters linked to a subset of sputum omics. *J Allergy Clin Immunol*. 2017;139(6):1797-807.
9. Baines KJ, Simpson JL, Wood LG, Scott RJ, Fibbens NL, Powell H, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. *J Allergy Clin Immunol*. 2014;133(4):997-1007.
10. Berthon BS, Gibson PG, Wood LG, MacDonald-Wicks LK, Baines KJ. A sputum gene expression signature predicts oral corticosteroid response in asthma. *Eur Respir J*. 2017;49(6).
11. Kuo CS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, et al. A Transcriptome-driven Analysis of Epithelial Brushings and Bronchial Biopsies to Define Asthma Phenotypes in U-BIOPRED. *Am J Respir Crit Care Med*. 2017;195(4):443-55.
12. Tsai YH, Parker JS, Yang IV, Kelada SNP. Meta-analysis of airway epithelium gene expression in asthma. *Eur Respir J*. 2018;51(5).
13. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, et al. Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc Natl Acad Sci U S A*. 2007;104(40):15858-63.
14. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, et al. T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med*. 2009;180(5):388-95.

15. Peters MC, Mekonnen ZK, Yuan S, Bhakta NR, Woodruff PG, Fahy JV. Measures of gene expression in sputum cells can identify TH2-high and TH2-low subtypes of asthma. *J Allergy Clin Immunol*. 2014;133(2):388-94.
16. Takayama G, Arima K, Kanaji T, Toda S, Tanaka H, Shoji S, et al. Periostin: a novel component of subepithelial fibrosis of bronchial asthma downstream of IL-4 and IL-13 signals. *J Allergy Clin Immunol*. 2006;118(1):98-104.
17. Fricker M, Heaney LG, Upham JW. Can biomarkers help us hit targets in difficult-to-treat asthma? *Respirology*. 2017;22(3):430-42.
18. Simpson JL, Yang IA, Upham JW, Reynolds PN, Hodge S, James AL, et al. Periostin levels and eosinophilic inflammation in poorly-controlled asthma. *BMC Pulm Med*. 2016;16(1):67.
19. Jia G, Erickson RW, Choy DF, Mosesova S, Wu LC, Solberg OD, et al. Periostin is a systemic biomarker of eosinophilic airway inflammation in asthmatic patients. *J Allergy Clin Immunol*. 2012;130(3):647-54 e10.
20. Wood LG, Garg ML, Gibson PG. A high-fat challenge increases airway inflammation and impairs bronchodilator recovery in asthma. *J Allergy Clin Immunol*. 2011;127(5):1133-40.
21. Wood LG, Garg ML, Smart JM, Scott HA, Barker D, Gibson PG. Manipulating antioxidant intake in asthma: a randomized controlled trial. *Am J Clin Nutr*. 2012;96(3):534-43.
22. Scott HA, Gibson PG, Garg ML, Pretto JJ, Morgan PJ, Callister R, et al. Dietary restriction and exercise improve airway inflammation and clinical outcomes in overweight and obese asthma: a randomized trial. *Clin Exp Allergy*. 2013;43(1):36-49.

23. Simpson JL, Powell H, Boyle MJ, Scott RJ, Gibson PG. Clarithromycin targets neutrophilic airway inflammation in refractory asthma. *Am J Respir Crit Care Med*. 2008;177(2):148-55.
24. Simpson JL, Baines KJ, Ryan N, Gibson PG. Neutrophilic asthma is characterised by increased rhinosinusitis with sleep disturbance and GERD. *Asian Pac J Allergy Immunol*. 2014;32(1):66-74.
25. Gibson PG, Yang IA, Upham JW, Reynolds PN, Hodge S, James AL, et al. Effect of azithromycin on asthma exacerbations and quality of life in adults with persistent uncontrolled asthma (AMAZES): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2017;390(10095):659-68.
26. Cordova-Rivera L, Gibson PG, Gardiner PA, Powell H, McDonald VM. Physical Activity and Exercise Capacity in Severe Asthma: Key Clinical Associations. *J Allergy Clin Immunol Pract*. 2018;6(3):814-22.
27. Maestre-Batlle D, Pena OM, Hirota JA, Gunawan E, Rider CF, Sutherland D, et al. Novel flow cytometry approach to identify bronchial epithelial cells from healthy human airways. *Sci Rep*. 2017;7:42214.
28. Hoessling Hv. *Das Sputum*: Springer-Verlag Berlin Heidelberg; 1921.
29. Naylor B. The shedding of the mucosa of the bronchial tree in asthma. *Thorax*. 1962;17:69-72.
30. Demarche S, Schleich F, Henket M, Paulus V, Van Hees T, Louis R. Detailed analysis of sputum and systemic inflammation in asthma phenotypes: are paucigranulocytic asthmatics really non-inflammatory? *BMC Pulm Med*. 2016;16:46.
31. Epithelial Desquamation in Asthma. *Am J Respir Crit Care Med*. 2001;164(10):1997.

32. Xiao C, Puddicombe SM, Field S, Haywood J, Broughton-Head V, Puxeddu I, et al. Defective epithelial barrier function in asthma. *J Allergy Clin Immunol*. 2011;128(3):549-56 e1-12.
33. Modena BD, Bleecker ER, Busse WW, Erzurum SC, Gaston BM, Jarjour NN, et al. Gene Expression Correlated with Severe Asthma Characteristics Reveals Heterogeneous Mechanisms of Severe Disease. *Am J Respir Crit Care Med*. 2017;195(11):1449-63.
34. Vermeer PD, Denker J, Estin M, Moninger TO, Keshavjee S, Karp P, et al. MMP9 modulates tight junction integrity and cell viability in human airway epithelia. *Am J Physiol Lung Cell Mol Physiol*. 2009;296(5):L751-62.
35. Dorscheid DR, Wojcik KR, Sun S, Marroquin B, White SR. Apoptosis of airway epithelial cells induced by corticosteroids. *Am J Respir Crit Care Med*. 2001;164(10 Pt 1):1939-47.
36. Choksi SP, Lauter G, Swoboda P, Roy S. Switching on cilia: transcriptional networks regulating ciliogenesis. *Development*. 2014;141(7):1427-41.
37. Hackett NR, Butler MW, Shaykhiev R, Salit J, Omberg L, Rodriguez-Flores JL, et al. RNA-Seq quantification of the human small airway epithelium transcriptome. *BMC Genomics*. 2012;13:82.
38. Geremek M, Bruinenberg M, Zietkiewicz E, Pogorzelski A, Witt M, Wijmenga C. Gene expression studies in cells from primary ciliary dyskinesia patients identify 208 potential ciliary genes. *Hum Genet*. 2011;129(3):283-93.
39. Thomas MP, Liu X, Whangbo J, McCrossan G, Sanborn KB, Basar E, et al. Apoptosis Triggers Specific, Rapid, and Global mRNA Decay with 3' Uridylated Intermediates Degraded by DIS3L2. *Cell Rep*. 2015;11(7):1079-89.

40. Bhakta NR, Christenson SA, Nerella S, Solberg OD, Nguyen CP, Choy DF, et al. IFN-stimulated Gene Expression, Type 2 Inflammation, and Endoplasmic Reticulum Stress in Asthma. *Am J Respir Crit Care Med*. 2018;197(3):313-24.
41. Fricker M, Gibson PG, Powell H, Simpson JL, Yang IA, Upham JW, et al. A sputum 6 gene signature predicts future exacerbations of poorly controlled asthma. *J Allergy Clin Immunol*. 2019.

Figure legends

Figure 1. Design of study including cohort details.

Figure 2. Flow cytometry and microscope measurement of sputum columnar epithelial cells and cell viability. (A) Micrograph of a cluster of sputum cells identified as columnar epithelial cells (CEPs) on the basis of morphology. CEPs identified by arrows. **(B)** Correlation of sputum CEP proportion measured in the same sample using two separate methods – traditional microscope differential cell count (DCC) and flow cytometry (FC) (n = 18, Spearman's, $r = 0.85$, $p < 0.001$). **(C)** Bland-Altman plot of sputum CEP quantification comparing DCC and FC methods. **(D)** Flow cytometry measurement of the percentage of sputum CEPs that were viable based on exclusion of 7-AAD (n = 8, mean, SD). **(E)** Correlation analysis of % sputum CEP viability with sputum CEP proportion (n = 8, Spearman's, $r = 0.01$, $p = 0.98$). **(F)** Correlation analysis of % sputum CEP viability with sputum CEP absolute count (n = 8, Spearman's, $r = 0.16$, $p = 0.7$).

Figure 3. qPCR validation of CEP high sputum gene signature and biomarker potential. Comparison of SCGB3A1 **(A)**, LDLRAD1 **(B)**, FOXJ1 **(C)**, DNALI1 **(D)**, CFAP53 **(E)** and CFAP157 **(F)** mRNA expression in CEP high (n = 29) vs low (n = 32) sputum samples.

Target mRNA is normalized relative to expression of the housekeeper gene β -actin for each sample (data displayed as mean $\pm$ SEM, Mann-Whitney test, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$). **(G)** ROC analysis of predicted values generated from multivariate logistic regression for SCGB3A1, LDLRAD1, FOXJ1, DNALI1, CFAP53 and CFAP157 (Epithelial sig, AUC 84.4%, $p < 0.0001$) expression combined with CEP high vs low status as the dependent variable. ROC curve for CEP high vs low status comparison using DNALI1 values only also shown (DNALI1, AUC 83.1%, $p < 0.0001$).

Figure 4. Type 2 epithelial gene signature expression and periostin protein measures in sputum are elevated only in those asthma patients with both high CEP and high eosinophil status. Comparison of *CLCA1* **(A)** and *SerpinB2* **(B)** mRNA expression in CEP high (n = 29) vs low (n = 32) sputum samples. Comparison of *CLCA1* **(C)** and *SerpinB2* **(D)** mRNA expression in CEP low Eosinophil (Eo) low (n = 13), CEP high Eo low (n = 18), CEP low Eo high (n = 18) and CEP high Eo high (n = 12) sputum samples. In **(A-D)** target mRNA is normalized relative to expression of the housekeeper gene β -actin for each sample. ELISA measurement sputum supernatant periostin protein **(E&G)** in CEP high (n = 23) and low (n = 36) **(E)**, and CEP low Eo low (n = 23), CEP high Eo low (n = 13), CEP low Eo high (n = 13) and CEP high Eo high (n = 10) **(G)** samples. ELISA measurement of serum periostin protein **(F&H)** in CEP high (n = 26) and low (n = 55) **(F)**, and CEP low Eo low (n = 32), CEP high Eo low (n = 17), CEP low Eo high (n = 23) and CEP high Eo high (n = 9) **(H)** samples. For **(A-D)** data displayed as mean $\pm$ SD, **(A&B, E&F)** Mann-Whitney test, **(C&D, G&H)** Kruskal-Wallis with multiple comparisons, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$). ROC analysis of sputum supernatant periostin protein concentration with CEP high vs CEP low status as dependent variables **(I)** – AUC 63.1%, $p = 0.09$ and CEP high Eo high vs all other patients as dependent variables **(J)** – AUC 83.7%, $p = 0.0009$.

Tables

Table 1 – Subject characteristics and sputum differential cell counts

	Control (n=138)	Asthma (n=901)	P value
Age (y) [†]	50.17 (16.59), (n=138)	54.23 (15.55), (n=901)	.005*
Male [†]	57 (41.3), (n=138)	381 (42.29), (n=901)	.828
BMI (kg/m ²) [‡]	26.54 (4.56), (n=127)	30.31 (7.04), (n=823)	<.001*
Smoker (current or ex) [†]	62 (48.44), (n=128)	338 (41.12), (n=882)	<.001*
Atopy [†]	44 (51.63), (n=86)	599 (72.78), (n=823)	<.001*
FENO (ppb) [§]	17.1 (9.7-23.73), (n=113)	21.9 (13.85, 40.11), (n=549)	<.001*
Pre-β2 FEV1 (%) [§]	93.38 (83.55-105.6), (n=136)	78.46 (62.96-93), (n=895)	<.001*
Pre-β2 FEV1/FVC (%) [§]	78.57 (73.45-83.76), (n=134)	70.9 (62.55-78.15), (n=893)	<.001*
Severe asthma [†]		355 (39.75), (n=893)	
ACQ6 score [‡]		1.52 (0.99), (n=821)	
ICS treated [†]		763 (85.06), (n=900)	
AHR [†]		349 (49.57), (n=704)	
Sputum TCC (×10 ⁶ /mL) [§]	3.11 (1.62-5.18), (n=122)	3.78 (2.07-7.11) (n=846)	.001*
Neutrophils (×10 ⁴ /mL) [§]	72.71 (27.75-209.05)	119.66 (45.56-307.44)	.001*
Eosinophils (×10 ⁴ /mL) [§]	1.72 (0.28-5.78)	7.73 (1.49-30.59)	<.001*
Macrophages (×10 ⁴ /mL) [§]	152.33 (93.1-271.01)	164.84 (92.90-265.06)	.961
Lymphocytes (×10 ⁴ /mL) [§]	4.64 (0.81-13.62)	2.72 (0.51-7.71)	.011*
CEPs (×10 ⁴ /mL) [§]	3.31 (0.96-9.95)	5.47 (1.38-18.10)	.002*
Neutrophils (%) [§]	29.15 (13-47.95), (n=138)	35.5 (19.25-57), (n=901)	.003*
Eosinophils (%) [§]	0.5 (0-1.25)	1.75 (0.5-7.75)	<.001*
Macrophages (%) [§]	61.38 (43.42-78.63)	49.25 (30.88-65.25)	<.001*
Lymphocytes (%) [§]	1.25 (0.25-3)	0.75 (0.25-1.75)	.002*
CEPs (%) [§]	1.25 (0.25-3.06)	1.75 (0.5-4.75)	.057

*denotes a statistically significant difference between healthy and asthma

Data are presented as numbers (percentages)[†], mean (SD)[‡] or medians (25th-75th percentiles)[§].

Abbreviations: BMI (Body Mass Index), FENO (Fractional Exhaled Nitric Oxide), AHR (Airways Hyperresponsiveness), TCC (Total Cell Count), CEPs (Columnar Epithelial Cells).

Table 2 – Univariate and multivariate linear regression analysis for factors associated with sputum CEP percentage in patients with asthma

Factor	Univariate analysis Coefficient (95% CI)	<i>P</i> value	Multivariate analysis Coefficient (95% CI)	<i>P</i> value
Age (y)	-0.003 (-0.007-0.001)	0.205	0.001 (-0.003-0.005)	0.666
Gender	0.196 (0.071-0.332)	0.002*	0.279 (0.153-0.406)	<0.001*
BMI (kg/m ²)	0.003 (-0.006-0.012)	0.538		
Smoking packyears	-0.005 (-0.010--0.001)	0.016*	-0.005 (-0.010--0.001)	0.014*
Atopy	0.002 (-0.143-0.147)	0.979		
FENO (ppb)	0.001 (-0.001-0.004)	0.335		
FEV1 (%)	0.004 (0.001-0.007)	0.009*	0.001 (-0.004-0.005)	0.671
FEV1/FVC (%)	0.007 (0.002-0.013)	0.008*	0.002 (-0.006-0.011)	0.560
Poorly-controlled	0.133 (0.001-0.265)	0.049*		
AHR	-0.080 (-0.221-0.060)	0.262		
No ICS use	0 [reference]			
Low dose	-0.114 (-0.347-0.119)	0.337		
Medium dose	0.007 (-0.192-0.205)	0.948		
High dose	0.165 (-0.023-0.353)	0.085		
Severe asthma	0.192 (0.065-0.353)	0.003*	0.285 (0.160-0.411)	<0.001*
Sputum TCC (×10 ⁶ /mL)	-0.033 (-0.043--0.023)	<0.001*	0.053 (0.026-0.081)	<0.001*
Neutrophils (×10 ⁴ /mL)	-0.001 (-0.001-0.000)	<0.001*	-0.001 (-0.001--0.001)	<0.001*
Eosinophils (×10 ⁴ /mL)	0.000 (-0.001-0.000)	0.146	-0.001 (-0.001-0.000)	0.071
Macrophages (×10 ⁴ /mL)	0.000 (0.000-0.000)	0.962		
Lymphocytes (×10 ⁴ /mL)	-0.007 (-0.014--0.001)	0.034*	0.000 (-0.007-0.007)	0.928

*denotes a statistically significant correlation with sputum CEP percentage

Abbreviations: BMI (Body Mass Index), FENO (Fractional Exhaled Nitric Oxide), AHR (Airways Hyperresponsiveness), ICS (Inhaled Corticosteroid), TCC (Total Cell Count).

Table 3 – Clinical and inflammatory characteristics of CEP % high asthma

	CEP % low (n=805 unless stated)	CEP % high (n=96 unless stated)	P value
Age (y) [‡]	54.31 (15.55)	53.56 (15.64)	0.658
Male [†]	334 (41.49)	47 (48.96)	0.162
BMI (kg/m ²) [‡]	30.20 (6.99), (n=745)	31.35 (7.45), (n=78)	0.171
Smoking pack/years [§]	0 (0-3.0), (n=785)	0 (0-0.5), (n=89)	0.063
Atopy [†]	535 (72.89), (n=734)	64 (71.91), (n=89)	0.864
FENO (ppb) [§]	21.83 (13.60-39.91), (n=495)	23.29 (14.53-40.8), (n=55)	0.892
Pre-β2 FEV1 (%) [§]	77.78 (62.96-93.0), (n=799)	82.68 (61.75-95.0)	0.163
Pre-β2 FEV1/FVC (%) [§]	70.27 (62.28-77.89), (n=797)	74.13 (63.0-81.11)	0.013*
Poorly controlled [†]	321 (43.79), (n=733)	43 (48.86) (n=88)	0.366
AHR [†]	318 (50.4), (n=631)	31 (42.47), (n=73)	0.390
ICS use:	n=796	n=96	
No ICS [†]	126 (15.83)	8 (8.33)	0.052
Low dose [†]	112 (14.07)	11 (11.46)	0.483
Medium dose [†]	232 (29.15)	27 (28.13)	0.835
High dose [†]	326 (40.95)	50 (52.08)	0.037*
Severe asthma [†]	306 (38.39), (n=797)	49 (51.04), (n=96)	0.017*
Sputum cell count:	n=760	n=86	
Sputum TCC (×10 ⁶ /mL) [§]	3.83 (2.16-7.29)	2.94 (1.53-5.42)	0.004*
Neutrophils (×10 ⁴ /mL) [§]	138.72 (52.02-328.64)	48.0 (16.2-91.28)	<0.001*
Eosinophils (×10 ⁴ /mL) [§]	8.47 (1.62-30.88)	2.97 (0.54-28.13)	0.010*
Macrophages (×10 ⁴ /mL) [§]	165.68 (96.12-269.28)	143.88 (73.53-248.21)	0.109
Lymphocytes (×10 ⁴ /mL) [§]	2.91 (0.59-8.15)	1.91 (0-5.09)	0.018*
CEPs (×10 ⁴ /mL) [§]	4.41 (0.99-12.56)	57.62 (26.06-95.33)	<0.001*
Neutrophils (%) [§]	37.75 (22.25-59.0)	17.31 (6.81-34)	<0.001*
Eosinophils (%) [§]	1.94 (0.5-7.75)	1 (0.25-7.58)	0.208
Macrophages (%) [§]	49.25 (29.63-66.0)	49.5 (39.0-64.06)	0.427
Lymphocytes (%) [§]	0.75 (0.25-1.75)	0.75 (0-1.5)	0.396
CEPs (%) [§]	1.5 (0.25-3.5)	17.91 (13.35-23)	<0.001*
Sputum inflammatory phenotype [†]	EA 287 (35.65) NA 131 (16.27) PGA 332 (41.24) MGA 55 (6.83)	37 (38.54) 2 (2.08) 55 (57.29) 2 (2.08)	0.577 <0.001* 0.003* 0.071

CEP % high defined by sputum CEP ≥11%

*denotes a statistically significant difference between low Cep percentage and high Cep percentage asthma.

Data are presented as numbers (percentages)[†], mean (SD)[‡] or medians (25th-75th percentiles)[§].

Abbreviations: BMI (Body Mass Index), FENO (Fractional Exhaled Nitric Oxide), AHR (Airways Hyperresponsiveness), ICS (Inhaled Corticosteroid), TCC (Total Cell Count), EA (Eosinophilic asthma), NA (Neutrophilic asthma), PGA (Paucigranulocytic asthma), MGA (Mixed granulocytic asthma).

COHORT 1.

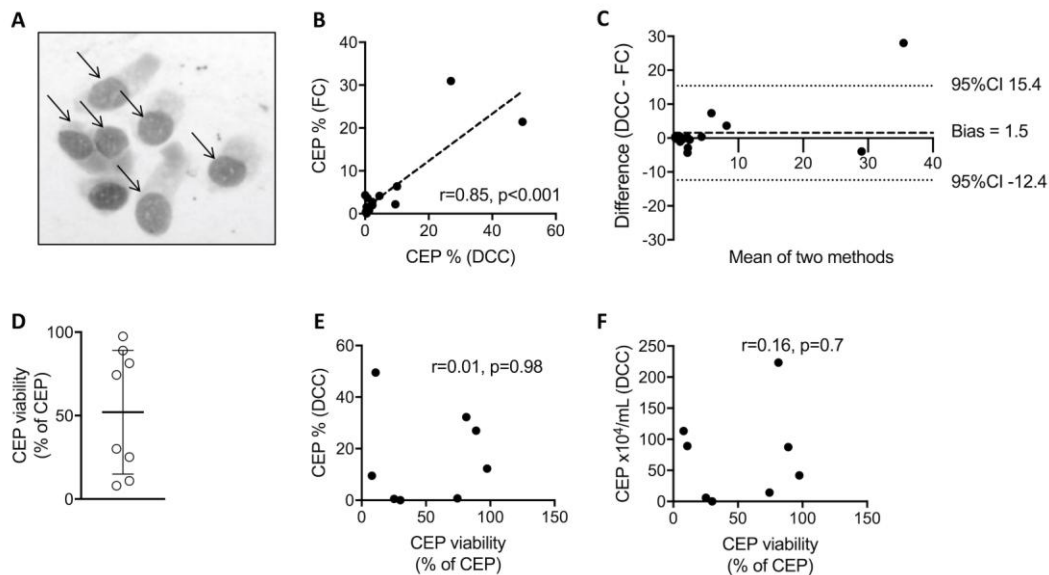
Retrospective sputum cohort, HMRI sputum database 2005-2015. 901 participants with asthma, 138 healthy controls. Includes 59 participants with asthma for sputum microarray analysis, 81 participants with asthma for Periostin ELISA in serum and sputum.

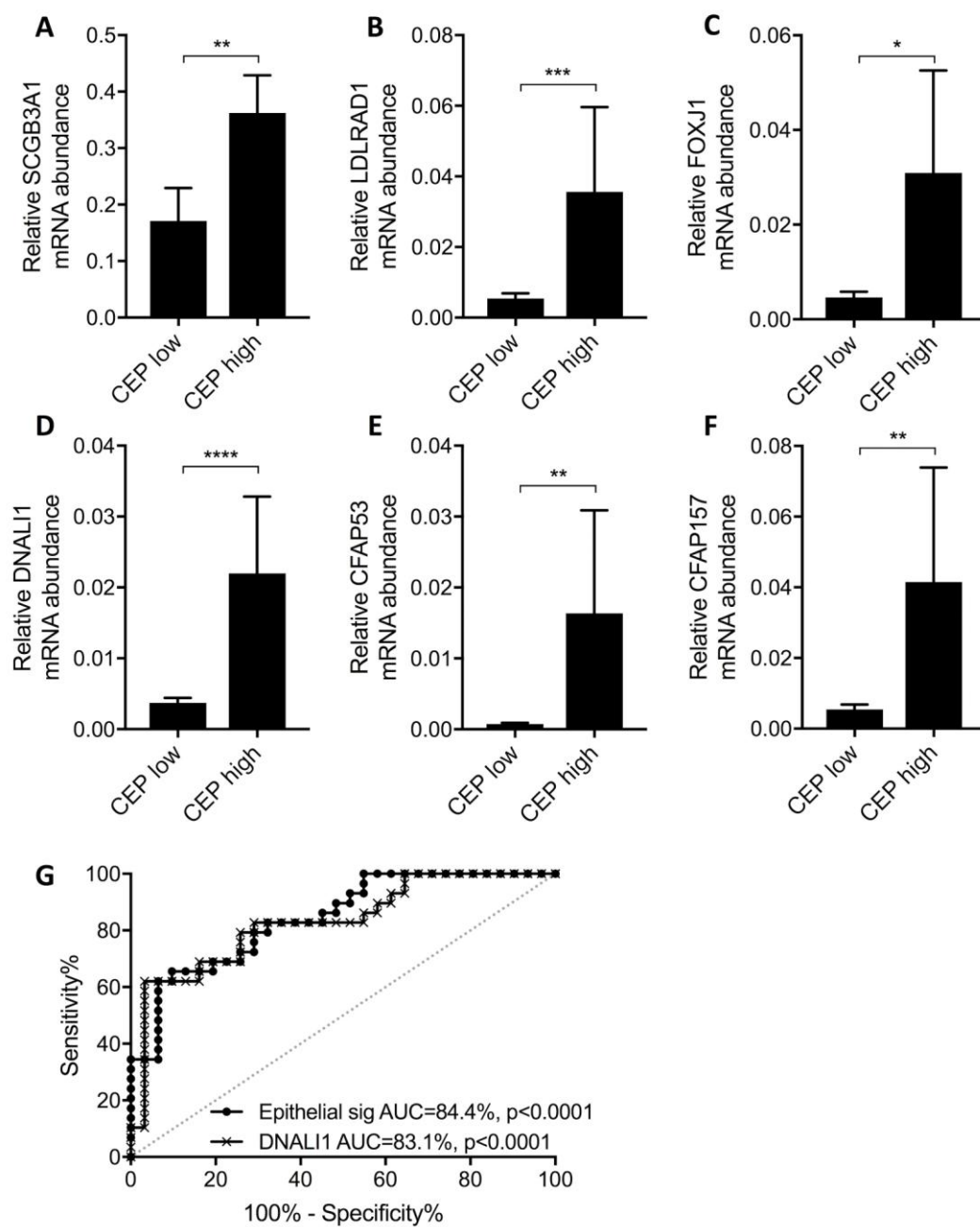
COHORT 2.

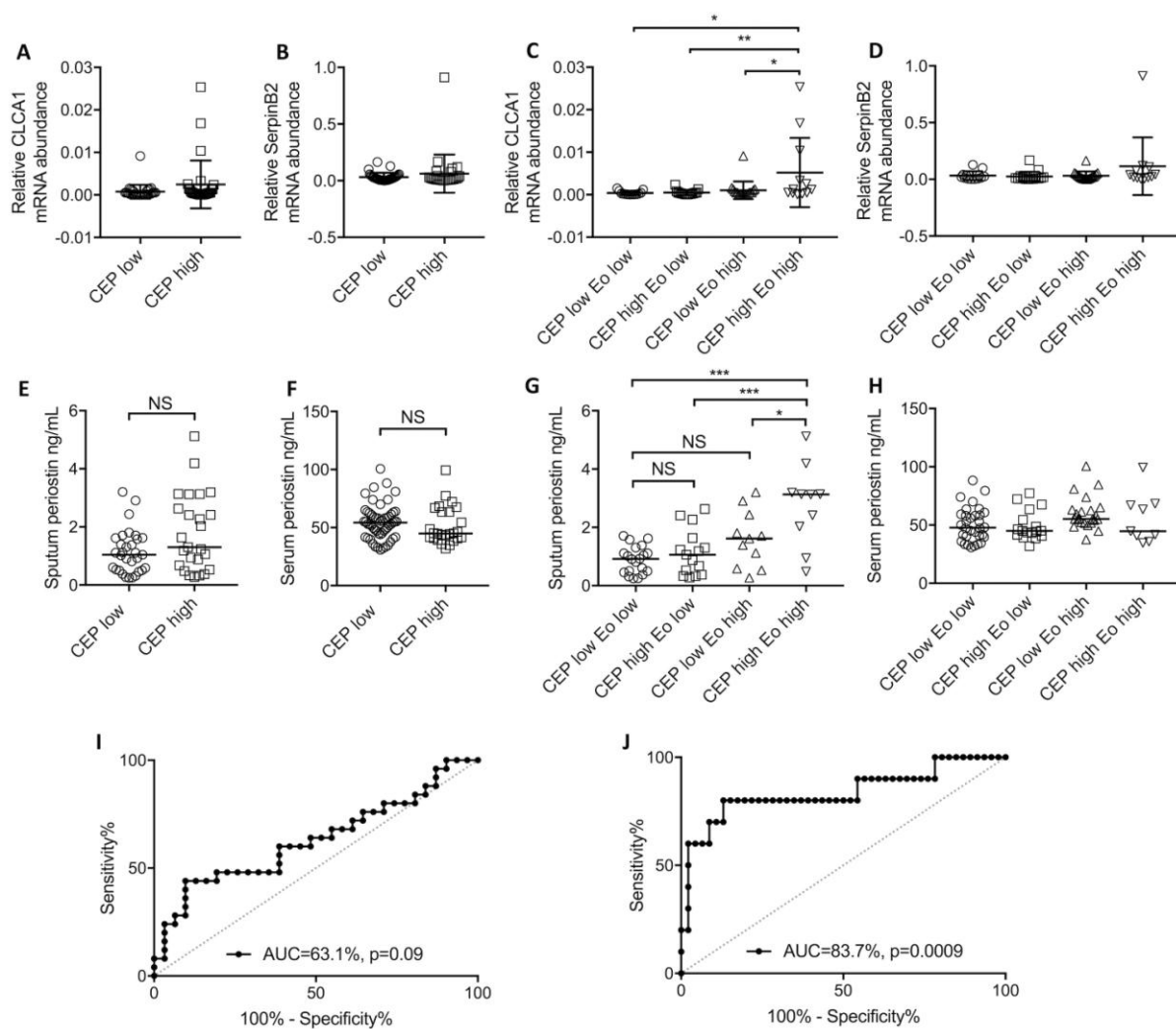
Epithelial gene signature qPCR validation and type 2 epithelial gene signature qPCR measurement. 60 participants with asthma.

COHORT 3.

Sputum flow cytometry measures. 17 participants with asthma.







Dysregulation of sputum columnar epithelial cells and products in distinct asthma phenotypes.

Authors: Ling Qin, Peter G. Gibson, Jodie L. Simpson, Katherine J. Baines, Vanessa M. McDonald, Lisa G. Wood, Heather Powell, Michael Fricker

ONLINE SUPPLEMENTAL INFORMATION

Methods

Microarray analysis and sputum qPCR analysis populations

We conducted 2 studies: a discovery study (n=59 patients with asthma who were part of cohort 1) and a clinical validation study (n=60 patients with asthma, cohort 2). Our microarray data and microarray study population details were previously published [1]. For microarray CEP high patients were defined as having CEP proportion $\geq 11\%$ and/or CEP absolute count $\geq 18.1 \times 10^4/\text{mL}$. For the validation study, adults with stable asthma were recruited from John Hunter Hospital, Newcastle, Australia. Asthma was diagnosed according to American Thoracic Society guidelines on the basis of current (past 12 months) episodic respiratory symptoms, doctor's diagnosis (ever), and demonstrated evidence of AHR to hypertonic saline. Exclusion criteria included recent (past month) respiratory tract infection, recent asthma exacerbation, recent unstable asthma or change in maintenance therapy, and current smoking. All participants gave written informed consent before their inclusion in the study. For qPCR validation CEP high patients were defined as having CEP proportion $\geq 11\%$. For the purposes of qPCR validation of mRNAs increased in CEP high asthma, a roughly equal proportion of CEP high (n = 29) vs CEP low (n = 31) sputum samples were selected for analysis. Patient characteristics of the qPCR validation study are provided in Supplemental Table 1.

Definitions and clinical variables

Severe asthma was defined according to the American Thoracic Society/European Respiratory Society (ATS/ERS) guidelines in 2014, as asthma that requires treatment with high-dose inhaled corticosteroids (ICS) plus a second controller and/or systemic corticosteroids to prevent it from becoming "uncontrolled" or that remains "uncontrolled"

despite this therapy [2]. Asthma control was assessed using the asthma control questionnaire (ACQ) according to GINA 2017 guidelines. Controlled status was defined as poor (ACQ > 1.5) or not poor (ACQ $\leq$ 1.5). Inhaled corticosteroid dose was defined as low, medium or high according to GINA 2017 guidelines[3]. The clinical variables analyzed were: gender (male/female), age, BMI (weight (kg)/height² (m)), smoking index (pack/years, non-smokers were set as 0), FENO (ppb), atopy (skin prick test, positive/negative or positive specific IgE to one or more common aeroallergens), FEV₁%predicted (Forced expiratory volume in 1 second), FEV₁/FVC (%) (forced vital capacity), AHR (high dose saline PD15 <15 mg/mL, y/n) (airways hyperresponsiveness), severe asthma (y/n), inhaled corticosteroid (ICS) use (no/low/medium/high).

Sputum induction and processing

Sputum induction with hypertonic saline (4.5%) was performed in participants whose FEV₁ was $\geq$ 1L using our previously described methods [4, 5]. In those with FEV₁ <1L, 0.9% saline was used. For gene expression, Buffer RLT (Qiagen, Hilden, Germany) was immediately added to 100 μ L of selected sputum and stored at -80 °C until RNA extraction. For inflammatory cell counts and sputum flow cytometry, selected sputum was dispersed using dithiothreitol, and total cell count and viability were performed. Cytospins were prepared, stained (May-Grunwald–Giemsa) and a differential cell count obtained from 400 non-squamous cells. Sputum cell suspension supernatant was stored at -80 °C for ELISA.

Sputum flow cytometry study

Adults with stable asthma were recruited from John Hunter Hospital, Newcastle, Australia. Asthma was diagnosed according to American Thoracic Society guidelines on the basis of current (past 12 months) episodic respiratory symptoms, doctor's diagnosis (ever), and

demonstrated evidence of AHR to hypertonic saline. Exclusion criteria included recent (past month) respiratory tract infection, recent asthma exacerbation, recent unstable asthma or change in maintenance therapy, and current smoking. All participants gave written informed consent before their inclusion in the study. Sputum flow cytometry was performed on freshly isolated sputum cell suspensions [6]. Dispersed sputum cells were resuspended for subsequent steps in flow cytometry buffer (PBS containing 1% FCS, 0.01% sodium azide and 2 mM ethylenediaminetetraacetic acid [EDTA]) at a density of 1×10^6 / mL. All steps were carried out at 4 °C. Cell suspension was incubated with human Fc Block (BD Biosciences) following manufacturer's instructions, washed in flow cytometry buffer twice, then labelled with the following antibodies in flow cytometry buffer (CD45-APC, EpCAM-BV605 – both BD Biosciences). Cells were washed again and resuspended in flow cytometry buffer containing the viability dye 7-aminoactinomycin D (7-AAD) at a final concentration of 0.5 µg/mL. Data acquisition was performed on a FACS Aria III flow cytometer (BD Bioscience). Data analysis was performed using Flowjo software (BD Bioscience). Patient characteristics of the sputum flow cytometry cohort are provided in Supplemental Table 1.

Sputum RNA extraction

For qPCR and microarray studies, 100 µL of selected sputum was stored in buffer RLT (Qiagen, Hilden, Germany) at -80 °C. RNA was extracted from induced sputum samples using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. RNA was quantitated using the Quanti-iT RiboGreen RNA Quantitation Assay Kit (Molecular Probes Inc., Invitrogen, Eugene, Oregon) as per manufacturer's instructions.

Sputum microarray analysis

Sputum microarray analysis was performed on previously published sputum microarray data from 59 patients with stable asthma[1]. A total of 500 ng of RNA was reverse transcribed into cRNA and biotin-uridine-5'-triphosphate labeled by using the Illumina TotalPrep RNA Amplification Kit (Ambion, Austin, Texas). A total of 750 ng cRNA was hybridized to the Illumina Sentrix HumanRef-8 Version 2 Expression BeadChips and scanned by using the Illumina Bead Station and captured using BeadScan 3.5.11 (Illumina, San Diego, California). Samples and gene profiling results were included in the analysis if the sample was of suitable purity (OD 260/280 1.7-2.1) and was successfully amplified (sufficient cRNA generated) and hybridized (95th percentile of fluorescence score >500) and the data passed quality controls in GeneSpring GX (correlation coefficients and principle component analysis plots; Agilent Technologies, Santa Clara, California).

Microarray data were exported by using Genome Studio (Illumina) and analyzed by using GeneSpring GX11 (Agilent Technologies). Data were log-transformed, normalized, and baseline-converted to the median of all samples. We divided the 59 asthma sputum samples analysed by microarray into two groups: high CEP (n=5) and low CEP (n=54). Data were filtered, and only genes flagged as present (<0.05 detection P value) in all 59 samples were included in the further analysis. Differential gene expression was determined using independent t-test ($P < 0.05$) and fold change >1.5. Gene ontology (GO) categories and pathways that were overrepresented in the differentially expressed gene lists were determined by using GeneSpring GX11 where an adjusted $P < 0.05$ was considered significant.

Candidate reporter transcripts for high CEP sputum samples were selected for the validation qPCR study by ranking and selection from the microarray-identified differentially expressed genes based on: P value of difference between the two groups, degree of fold change (>1.5 fold), abundance of mRNA expression in sputum (raw data>40), obvious

correlation with CEP percentage ($R>0.5$) and commercial availability of a suitable TaqMan prime/probe set.

Quantitative real-time PCR

Sputum qPCR samples were conducted on 60 samples from patients with stable asthma, using standard protocols [1, 7]. Sputum RNA (200 ng when possible, or maximal possible up to 200ng for samples with lower RNA concentrations) was reverse-transcribed to cDNA using the high capacity cDNA reverse transcription kit following manufacturer's instructions (Applied Biosystems, Foster City, California). Taqman qPCR primers and probes were combined with Taqman gene expression master mix as per the manufacturer's instructions in duplicate singleplex real-time PCRs (QuantStudio 6 Flex system; Applied Biosystems). cDNA samples were tested at various dilutions and highest dilutions which achieved Ct values below 30 for B-ACTIN were used for detection of epithelial gene signature target genes. Data are represented as and statistical analysis performed on $2^{-\Delta CT}$ values (ΔCT = cycle threshold of target - cycle threshold of β -actin).

Periostin ELISA

Periostin ELISA data, including patient recruitment and characteristics, was previously reported [8]. Venous blood samples were collected without anticoagulant. Periostin levels in serum were measured using the clinical trial version of the Elecsys® periostin immunoassay (Roche Diagnostics, Penzberg, Germany) intended for use on the cobas e 601 immunology analyzer, an automated electrochemiluminescence immunoassay based on the sandwich principle. Periostin levels in sputum supernatants were measured using the Genentech proprietary ELISA using the same antibody pair as the Elecsys® assay. The limit of detection

132 of this assay is 9.4 pg/mL. For periostin ELISA analysis CEP high patients were defined as
133 having CEP proportion $\geq 11\%$ and/or CEP absolute count $\geq 18.1 \times 10^4/\text{mL}$.

134

135

136

137 **Supplementary Tables**

138 Table E1 – Subject characteristics of the qPCR validation and sputum flow cytometry cohorts

	Cohort 2 Sputum qPCR cohort N=60	Cohort 3 Sputum flow cytometry cohort N=17
Age, (y) [‡]	58.87 (15.68)	59.47 (16.28)
Gender, M/F	27/33	9/8
BMI [‡]	31.41 (5.46)	29.07 (5.41)
Smoker (current or ex) [†]	14 (23.3%)	4 (23.5%)
Atopy [†]	28/40 (70.0%)	3 /4 (75.0%)
Mean ACQ score [‡]	1.57 (0.89)	0.98 (0.88)
<i>Asthma history</i>	N=60	N=12
Asthma duration, years [‡]	39.56 (20.34)	45.27 (16.90)
Maintenance ICS or ICS/LABA [†]	60 (100%)	15/15 (100%)
ICS dose, BDP equivalent mcg/day [§]	1000 (800, 2000)	1000 (800, 2000)
<i>Pre B2 spirometry[‡]</i>	N=60	N=12
FEV1 % predicted	77.04 (20.02)	71.63 (24.03)
FVC % predicted	84.41 (17.27)	82.63 (24.47)
FEV1/FVC	70.15 (11.98)	66.45 (10.92)
<i>Sputum cell counts</i>	N=60	N=17
Viability [‡]	63.31 (23.48)	79.79 (16.41)
Neutrophils%	26.75 (11.13, 56.50)	50.25 (37.50, 66.40)
Eosinophils%	4.0 (0.75, 14.25)	1.25 (0.60, 2.25)
Total cell count x10 ⁶ /ml	4.55 (2.12, 8.15)	8.46 (3.67, 12.0)

139 Data are presented as numbers (percentages)[†], mean (SD)[‡] or medians (25th-75th percentiles)[§].

140

Table E2 - Univariate and multivariate linear regression analysis for factors associated with sputum CEP count in patients with asthma (Multivariate analysis includes age, gender, smoking index, atopy, FENO, severe asthma, AHR, inflammatory phenotypes, *P* value=0.041)

Factor	Univariate analysis Coefficient (95% CI)	<i>P</i> value	Multivariate analysis Coefficients (95% CI)	<i>P</i> value
Age (y)	0.000 (-0.004-0.004)	0.885	-0.001 (-0.007-0.005)	0.751
Gender	0.246 (0.117-0.375)	<0.001*	0.312 (0.138-0.486)	<0.001*
BMI (kg/m ²)	0.004 (-0.006-0.014)	0.399		
Smoking packyears	-0.004 (-0.008-0.001)	0.083	-0.005 (-0.012-0.002)	0.192
Atopy	0.116 (-0.035-0.268)	0.131	0.070 (-0.118-0.257)	0.467
FENO (ppb)	0.002 (0.000-0.004)	0.095	0.002 (0.000-0.005)	0.079
FEV1 (%)	0.000 (-0.003-0.004)	0.834		
FEV1/FVC (%)	0.003 (-0.003-0.009)	0.309		
Poorly-controlled	0.209 (0.075-0.343)	0.002*		
AHR	-0.131 (-0.272-0.010)	0.068	-0.149 (-0.320-0.022)	0.087
No ICS use	0 [reference]			
Low dose	0.050 (-0.189-0.289)	0.680		
Medium dose	0.246 (0.040-0.452)	0.020*		
High dose	0.400 (0.205-0.584)	<0.001*		
Severe asthma	0.266 (0.136-0.397)	<0.001*	0.293 (0.119-0.466)	0.001*
<i>Sputum cell proportions</i>				
Neutrophils (%)	-0.008 (-0.011--0.006)	<0.001*	-0.007 (-0.011--0.004)	<0.001*
Eosinophils (%)	0.001 (-0.005-0.006)	0.819		
Macrophages (%)	0.002 (-0.001-0.004)	0.298		
Lymphocytes (%)	0.002 (-0.035-0.039)	0.911		

*denotes a statistically significant correlation with sputum CEP count

Abbreviations: BMI (Body Mass Index), FENO (Fractional Exhaled Nitric Oxide), AHR (Airways Hyperresponsiveness), ICS (Inhaled Corticosteroid).

150 Table E3 - Clinical and inflammatory characteristics of CEP count high asthma

	CEP count low (n=733 unless stated)	CEP count high (n=113 unless stated)	P value
Age (y) [‡]	54.34 (15.86)	55.09 (14.16)	0.636
Male [†]	305 (41.61)	60 (53.10)	0.025*
BMI (kg/m ²) [‡]	30.24 (6.89), (n=676)	30.96 (6.78), (n=93)	0.344
Smoking index (pack/years) [§]	0 (0-3.0), (n=718)	0 (0-0.8), (n=103)	0.214
Atopy [†]	476 (71.69), (n=664)	84 (80.0), (n=105)	0.144
FENO (ppb) [§]	23.2 (13.85-40.78), (n=458)	22.27 (16.98-40.18), (n=55)	0.461
Pre-β2 FEV1 (%) [§]	77.89 (63.93-93.28), (n=728)	77.77 (58.55-89.9), (n=112)	0.170
Pre-β2 FEV1/FVC (%) [§]	70.53 (62.30-77.73), (n=728)	69.84 (59.7-79.53), (n=112)	0.883
Poorly controlled [†]	288 (42.86), (n=672)	55 (54.46) (n=101)	0.029*
AHR [†]	298 (50.85), (n=586)	33 (45.21), (n=73)	0.001*
ICS use:	n=725	n=112	
No ICS [†]	114 (15.72)	6 (5.36)	0.004*
Low dose [†]	108 (14.90)	11 (9.82)	0.152
Medium dose [†]	209 (28.83)	33 (29.46)	0.890
High dose [†]	294 (40.55)	62 (55.36)	0.003*
Severe asthma [†]	277 (38.15), (n=726)	59 (52.68), (n=112)	0.004*
Sputum cell count:	n=733	n=113	
Sputum TCC (×10 ⁶ /mL) [§]	3.51 (1.98-6.3)	6.75 (3.87-10.62)	<0.001*
Neutrophils (×10 ⁴ /mL) [§]	115.52 (43.72-291.98)	147.47 (55.49-372.66)	0.090
Eosinophils (×10 ⁴ /mL) [§]	7.22 (1.41-26.84)	10.67 (2.13-74.81)	0.011*
Macrophages (×10 ⁴ /mL) [§]	153.70 (88.79-247.97)	275.4 (166.03-449.60)	<0.001*
Lymphocytes (×10 ⁴ /mL) [§]	2.7 (0.59-7.36)	3.33 (0-11.72)	0.416
CEPs (×10 ⁴ /mL) [§]	4.21 (0.9-11.41)	61.29 (43.74-94.01)	<0.001*
Neutrophils (%) [§]	36.5 (21.25-59.13)	29.75 (14.1-44.13)	<0.001*
Eosinophils (%) [§]	1.75 (0.5-7.5)	1.75 (0.5-10.75)	0.872
Macrophages (%) [§]	50.5 (30.18-66.38)	46.5 (33.13-58.13)	0.262
Lymphocytes (%) [§]	0.75 (0.25-2)	0.5 (0-1.5)	0.018*
CEPs (%) [§]	1.25 (0.25-3.25)	11.5 (5.83-20.5)	<0.001*
Sputum inflammatory phenotype [†]	EA 263 (35.88) NA 123 (16.78) PGA 300 (40.93) MGA 47 (6.41)	45 (39.82) 7 (6.19) 57 (50.44) 4 (3.54)	0.417 0.004* 0.057 0.232

CEP count high asthma was defined by having sputum CEP count $\geq 18.1 \times 10^4/\text{mL}$.

*denotes a statistically significant difference between CEP count low and CEP count high asthma.

Data are presented as numbers (percentages)[†], mean (SD)[‡] or medians (25th-75th percentiles)[§].

Abbreviations: BMI (Body Mass Index), FENO (Fractional Exhaled Nitric Oxide), AHR (Airways Hyperresponsiveness), ICS (Inhaled Corticosteroid), TCC (Total Cell Count), EA (Eosinophilic asthma), NA (Neutrophilic asthma), PGA (Paucigranulocytic asthma), MGA (Mixed granulocytic asthma).

Table E4 - Sputum inflammatory cell counts for CEP high and low asthma sputum samples used in microarray analysis

	CEP low (n=54)	CEP high (n=5)	P value
<i>Sputum cell count</i>			
TTC ($\times 10^6$ /mL)	3.38 (2.34-6.21)	2.25 (1.17-4.55)	.159
Neutrophils ($\times 10^4$ /mL)	146.43 (55.56-307.56)	70.38 (28.93-323.35)	.341
Eosinophils ($\times 10^4$ /mL)	12.84 (0.64-47.08)	0.52 (0-3.45)	.041*
Macrophages ($\times 10^4$ /mL)	102.89 (58.43-200.66)	62.73 (54.99-146.59)	.433
Lymphocytes ($\times 10^4$ /mL)	1.75 (0.05-3.95)	0 (0-3.74)	.262
CEPs ($\times 10^4$ /mL)	1.65 (0-4.71)	19.89 (17.06-27.97)	<0.001*
<i>Sputum cell proportions</i>			
Neutrophils (%)	50.25 (23.88-69.44)	46 (12.38-70.63)	.644
Eosinophils (%)	3 (0.25-12.38)	0.23 (0-2.5)	.065
Macrophages (%)	33.75 (22.19-56.75)	41 (19.93-71.25)	.625
Lymphocytes (%)	0.39 (0.19-1.5)	0 (0-1.48)	.359
CEPs (%)	0.5 (0, 2.13)	12.9 (11.01, 15.88)	<0.001*

Data are presented as medians (25th-75th percentiles), * denotes $p < 0.05$ between CEP high and low groups. Sputum samples were designated as being 'CEP high' on the basis of having a sputum CEP % value above the 90th percentile ($\geq 11\%$) OR a sputum CEP count value above the 90th percentile ($\geq 18.1 \times 10^4$ /mL) of reference healthy control values.

167 Table E5 – List of differentially expressed genes (p<0.05) identified in microarray
 168 comparison of CEP high and low asthma sputum samples.

169

Gene Symbol	CEP (n=54)	low	CEP high (n=5)	P value
RRAD	0.0943±0.5221		0.7125±0.7546	0.018
C1orf173	0.1295±0.4966		0.7725±0.9360	0.013
SCGB3A1	-0.2807±1.1463		1.2145±0.8859	0.006
LDLRAD1	0.1562±0.4881		0.9065±1.0212	0.005
CAPS variant1	0.2075±1.0476		1.5818±1.3629	0.008
DNAI1	0.1176±0.4868		0.7905±0.9532	0.009
FOXJ1	0.2642±0.8771		1.4806±1.2275	0.006
FZD4	0.1323±0.5296		0.8349±1.1578	0.014
GSTA2	0.2335±1.0692		1.6802±1.6872	0.008
CCDC113	0.1963±0.6153		0.9229±1.0357	0.021
WFDC6	0.1105±0.4713		0.6992±0.866	0.016
TPPP3	0.1164±0.6925		0.9031±0.9613	0.022
GSTA1	0.2355±1.0679		1.6773±1.7	0.008
LRRC50	0.2437±0.9738		1.3685±1.2934	0.019
C6orf165	0.1506±0.4695		0.7680±0.9951	0.015
TMEM190	0.2314±0.6821		1.1927±1.1924	0.007
RSPH1	0.2016±0.9740		1.3283±1.2450	0.019
SCGB1A1	0.0213±1.1417		1.3095±1.3034	0.02
CLDN3	0.1717±0.6684		0.9035±0.9856	0.028
TNFRSF19	0.2084±0.5624		0.8772±1.1268	0.024
CAPS variant2	0.1442±0.9478		1.5336±1.2750	0.003
LRRC46	0.1863±0.5876		0.8917±0.9299	0.018
PLUNC	0.1113±0.4837		0.7633±1.8209	0.042
C11orf70	0.1484±0.4708		0.7485±0.8715	0.015
HLA-DRB5	1.5808±3.1128		4.7256±4.3870	0.041
RAGE	0.0629±0.5171		0.7228±0.9219	0.014
AKAP14	0.1477±0.6049		0.7578±0.9235	0.044
TMEM146	0.064±0.6055		0.9321±0.8952	0.005
CTXN1	0.0997±0.8246		1.0187±1.0110	0.023
MS4A8B	0.2461±0.8652		1.2020±1.3312	0.028
GLDN	-0.314±1.0348		0.6834±0.5491	0.039
C2orf40	0.235±0.9221		1.4838±1.4056	0.008
LRRC23	0.1303±0.8317		1.1896±1.1798	0.11
LRRC23	0.1104±0.5338		0.7343±0.7522	0.019
ZMYND10	0.1991±0.9460		1.3740±1.1806	0.012
AQP3	-0.1091±0.4923		0.5302±0.2598	0.006
DNAH2AL	0.1985±0.6136		0.9598±1.1322	0.017
ALDH1L1	0.1408±0.6107		0.8009±0.8014	0.028
C9orf116	0.2052±0.8658		1.2823±1.1178	0.012
LRRC48	0.2131±0.5976		0.9495±0.9645	0.015
DNAL1	0.0792±0.5285		0.7857±0.9720	0.01
SLC27A2	0.1518±0.5429		0.7836±0.9020	0.022
SPA17	0.0775±0.5498		0.6981±0.7306	0.022
ARMC4	0.1522±0.5150		0.7492±0.8170	0.022
C9orf24	-0.0158±1.1137		1.2725±1.2662	0.017
TEKT1	0.1523±0.9084		1.2909±1.1722	0.011
TTC25	0.1711±0.5195		0.9168±0.9329	0.007
FBXO15	0.0681±0.5068		0.7276±0.8535	0.01
CGN	0.0715±0.4996		0.6980±0.8992	0.016
GRIN3B	0.1166±0.3986		0.8207±1.0633	0.003
DNAH9	0.1493±0.4943		0.7627±0.8900	0.017
FLJ23834	0.2655±0.8193		1.2429±1.3781	0.02
IGFBP5	0.1698±0.5393		0.7870±1.1087	0.031
CCNO	0.1185±0.5755		0.7285±0.7906	0.032
CFAP53	0.2336±0.5723		0.9952±0.9247	0.009
C13orf30	0.2093±0.8187		1.0579±1.0062	0.034
DYNLRB2	0.0674±0.9289		1.2255±1.1231	0.011
DNHD2	0.1746±0.5795		0.8597±1.1533	0.025
EFCAB1	0.1927±0.7577		1.2064±1.2859	0.009
CASC1	0.1387±0.5023		0.7535±0.8940	0.018
C20orf85	0.1539±0.9772		1.3403±1.1455	0.013
CFAP157	0.1807±0.9989		1.7557±1.8950	0.003
SLC2A5	0.3601±0.8042		-0.4456±0.4864	0.032

Sputum samples were designated as being 'CEP high' on the basis of having a sputum CEP % value above the 90th percentile ($\geq 11\%$) OR a sputum CEP count value above the 90th percentile ($\geq 18.1 \times 10^4/\text{mL}$) of reference healthy control values. Data presented are mean $\pm$ SD of normalized microarray values.

Table E6 – Significantly overrepresented gene ontology (GO) categories in genes upregulated in CEP high asthma sputum samples.

GO category	Adjusted P value
Cilium	0.02
Dynein complex	5.73×10^{-5}
Microtubule associated complex	0.008
Microtubule cytoskeleton	0.046
Microtubule motor activity	0.022
Motor activity	0.003
Single organism reproductive process	0.046

179 Table E7 – Selection of seven candidate sputum gene expression biomarkers of CEP high
180 status from microarray data for further qPCR validation studies

Gene symbol	Gene name	GenBank	CEP high vs low P value	Fold change CEP high vs low	R value of correlation with CEP %	
					All samples	High CEP samples only
SCGB3A1	Secretoglobin, family 3A, member 1	NM_052863.2	0.006	+2.82	0.443	0.891
LDLRAD1	Low density lipoprotein receptor class A domain containing 1	NM_001010978.2	0.005	+1.68	0.531	0.882
FOXJ1	Forkhead box J1	NM_001454.2	0.006	+2.32	0.505	0.893
DNALI1	Dynein, axonemal, light intermediate chain 1	NM_003462.3	0.01	+1.63	0.508	NS
TTC25	Tetratricopeptide repeat domain 25	NM_031421.2	0.007	+1.68	0.546	NS
CFAP53	Cilia and flagella associated protein 53	NM_145020.3	0.009	+1.70	0.501	NS
CFAP157	Cilia and flagella associated protein 157	NM_001012502.1	0.003	+2.98	0.544	NS

NS = correlation P value not significant (p>0.05)

184 Table E8 – Sputum cell counts for CEP high and low patients in cohort 2

	CEP low (n=29)	CEP high (n=31)	P value
<i>Sputum cell count</i>			
TCC ($\times 10^6/\text{mL}$)	7.65 (4.14-11.94)	2.79 (1.80-5.18)	<0.001*
Neutrophils ($\times 10^4/\text{mL}$)	312.19 (87.12-739.94)	50.51 (16.54-124.39)	<0.001*
Eosinophils ($\times 10^4/\text{mL}$)	20.36 (4.21-72.77)	5.66 (1.36-37.01)	.086
Macrophages ($\times 10^4/\text{mL}$)	187.72 (89.78-381.58)	88.80 (60.27-193.30)	.023*
Lymphocytes ($\times 10^4/\text{mL}$)	5.67 (1.08-11.34)	1.61 (0.00-7.87)	.215
CEPs ($\times 10^4/\text{mL}$)	6.93 (1.82-18.11)	63.18 (29.91-94.64)	<0.001*
<i>Sputum cell proportions</i>			
Neutrophils (%)	34.20 (19.76-72.00)	14.4 (5.50-36.00)	.001*
Eosinophils (%)	4.25 (0.75-12.75)	2.5 (0.38-19.50)	.784
Macrophages (%)	28.25 (16.25-65.25)	40.50 (28.00-57.38)	.391
Lymphocytes (%)	0.75 (0.25-1.75)	1.00 (0.00-2.82)	.714
CEPs (%)	1.25 (0.25-4.75)	17.25 (13.00-30.50)	<0.001*

185 Data are presented as medians (25th-75th percentiles), * denotes $p < 0.05$ between CEP high and low groups.
186 Sputum samples were designated as being ‘CEP high’ on the basis of having a sputum CEP % value above the
187 90th percentile ($\geq 11\%$) of reference healthy control values.
188

189

190

Table E9: Average Ct value of each gene and numbers of undetected samples (n=60)

Gene name	Average Ct value	Undetectable samples
B-ACTIN	27.72 (1.70)	0
SCGB3A1	30.92 (2.22)	0
LDLRAD1	35.61 (2.23)	3
FOXJ1	37.77 (2.39)	2
DNALI1	35.36 (2.30)	1
CFAP157	35.53 (2.46)	1
CFAP53	38.14 (2.02)	20 (10 in lowCep group)
CLCA1	38.95 (1.85)	33 (18 in lowCep group)
SERPINB2	33.45 (1.94)	0

The Ct values of undetectable samples were recorded as Ct value of 40 (the highest cycle number performed in PCR) in analysis. Data presented as mean (SD).

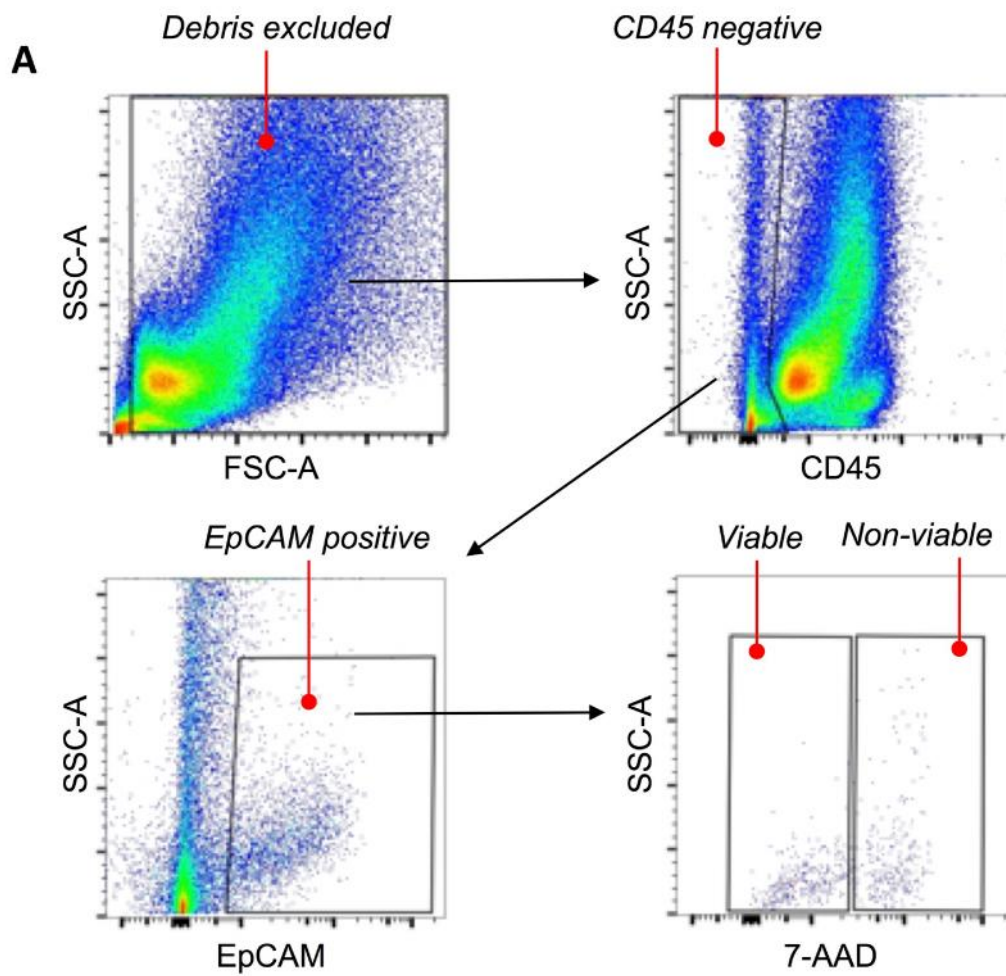
Table E10 – logistic regression and ROC analyses of epithelial signature gene expression and periostin protein as predictors of sputum CEP status.

Group comparison		Marker *	Logistic Regression				
			Constant	Coefficient	P value	Model value	P AUC (95%CI)
CEP high (N=29) vs CEP low (N=31)		SCGB3A1	-2.001	-2.439	0.169	0.001*	0.844
		LDLRAD1		64.430	0.364		(0.748,
		FOXJ1		-70.105	0.388		0.940)
		DNALI1		386.068	0.002*		P<0.0001*
		CFAP53		770.842	0.171		
		CFAP157		-102.251	0.078		
		DNALI1 alone	-1.587	248.379	0.002*	0.002*	0.831 (0.727, 0.935) P<0.0001*
CEP high (N=23) vs CEP low (N=36)		Periostin protein (sputum)	-1.612	0.798	0.008*	0.002*	0.682 (0.534, 0.831) P=0.019*
CEP high (N=26) vs CEP low (N=55)		Periostin protein (serum)	-0.053	-0.013	0.424	0.953	0.578 (0.438, 0.719) P=0.257
CEP high, Eo high (N=12) vs all others (N=48)		CLCA1	-1.913	309.772	0.042*	<0.0001*	0.742 (0.578, 0.908) P=0.01*
CEP high, Eo high (N=10) vs all others (N=49)		Periostin protein (sputum)	-4.746	1.673	0.001*	<0.0001*	0.878 (0.760, 0.995) P<0.0001*
CEP high, Eo high (N=9) vs all others (N=72)		Periostin protein (serum)	-2.484	0.007	0.740	0.053	0.506 (0.258, 0.730) P=0.952
CEP high, Eo high (N=12) vs CEP low, Eo low (N=13)		CLCA1	-1.138	1158.208	0.108	0.074	0.814 (0.647, 0.982) P=0.007*
CEP high, Eo high (N=10) vs CEP low, Eo low (N=23)		Periostin protein (sputum)	-4.671	2.346	0.003*	0.001*	0.922 (0.818, 1.000) P<0.0001*
CEP high, Eo high (N=9) vs CEP low, Eo low (N=32)		Periostin protein (serum)	-2.175	0.017	0.449	0.09	0.556 (0.329, 0.782) P=0.614

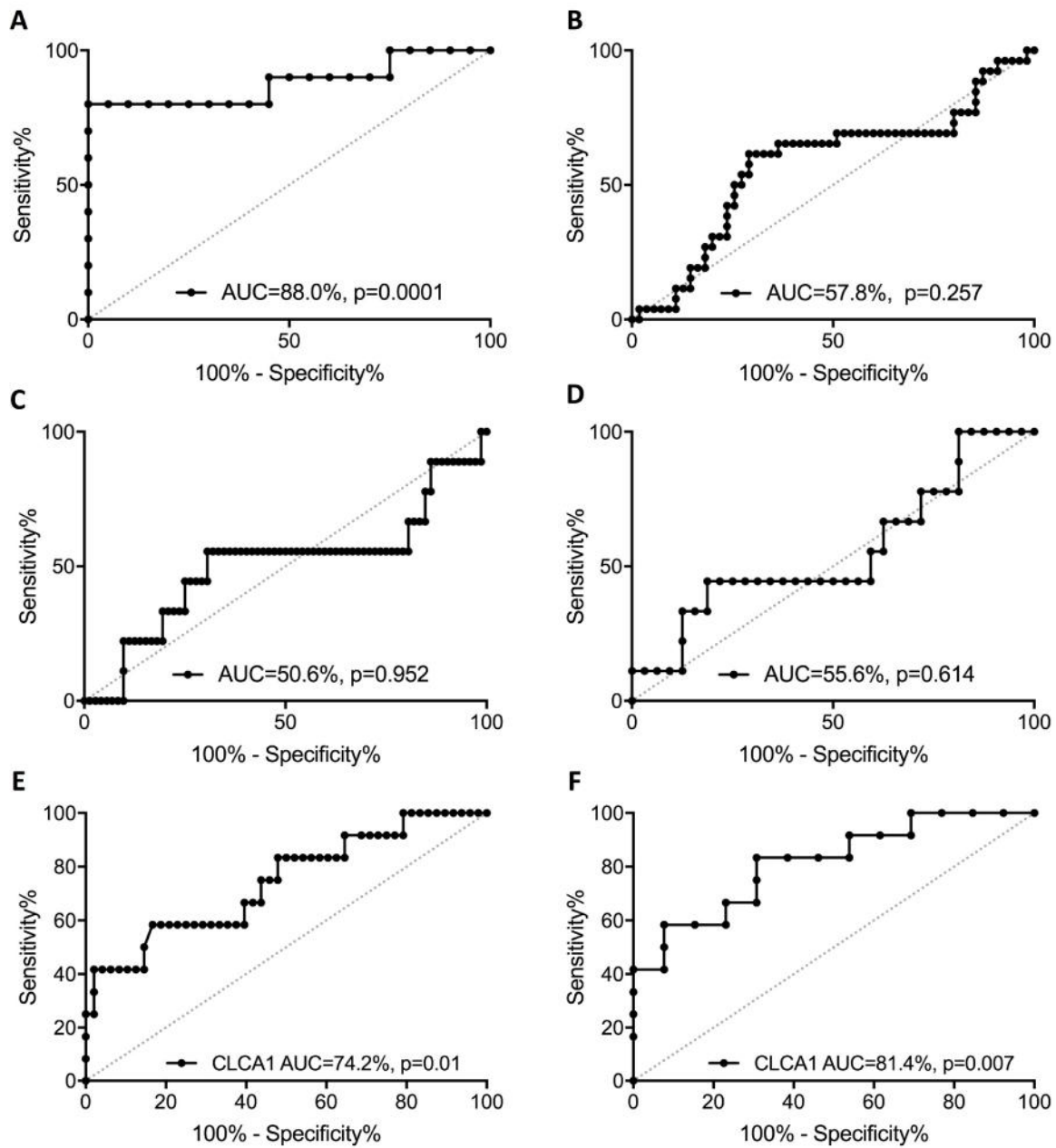
*Statistically significant (p<0.05)

†Markers are normalized to beta-actin mRNA expression (ΔCT)

202 Supplemental Figure E1. Flow cytometry gating strategy for the identification of CEPs
203 in sputum samples.



Supplemental Figure E2. ROC analysis of sputum and serum periostin protein and sputum CLCA1 mRNA for prediction of sputum CEP/Eosinophil categories.



Supplemental references

1. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. *J Allergy Clin Immunol*. 2011;127(1):153-60, 60 e1-9.
2. Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J*. 2014;43(2):343-73.
3. Global initiative for asthma. Global strategy for asthma management and prevention. 2017.
4. Gibson PG, Wlodarczyk JW, Hensley MJ, Gleeson M, Henry RL, Cripps AW, et al. Epidemiological association of airway inflammation with asthma symptoms and airway hyperresponsiveness in childhood. *Am J Respir Crit Care Med*. 1998;158(1):36-41.
5. Simpson JL, Scott R, Boyle MJ, Gibson PG. Inflammatory subtypes in asthma: assessment and identification using induced sputum. *Respirology*. 2006;11(1):54-61.
6. Brooks CR, van Dalen CJ, Hermans IF, Gibson PG, Simpson JL, Douwes J. Sputum basophils are increased in eosinophilic asthma compared with non-eosinophilic asthma phenotypes. *Allergy*. 2017;72(10):1583-6.
7. Baines KJ, Simpson JL, Wood LG, Scott RJ, Fibbens NL, Powell H, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. *J Allergy Clin Immunol*. 2014;133(4):997-1007.
8. Simpson JL, Yang IA, Upham JW, Reynolds PN, Hodge S, James AL, et al. Periostin levels and eosinophilic inflammation in poorly-controlled asthma. *BMC Pulm Med*. 2016;16(1):67.

Supplemental Figure Captions

Supplemental Figure E1. Flow cytometry gating strategy for the identification of CEPs in sputum samples. (A) Cellular debris was excluded on the basis of size, CD45 negative (non-leukocytes), EpCAM positive cells were selected as CEPs. In a subset of samples, CEP viability was determined based on 7-AAD staining, with 7-AAD negative events assigned viable status.

Supplemental Figure E2. ROC analysis of sputum and serum periostin protein and sputum CLCA1 mRNA for prediction of sputum CEP/Eosinophil categories. ROC analysis of **(A)** sputum periostin protein as a predictor of CEP high Eo high vs CEP low Eo low status (AUC 88.0%, $p = 0.0001$), **(B)** serum periostin protein as a predictor of CEP high vs low status (AUC 57.8%, $p = 0.257$), **(C)** serum periostin protein as a predictor of CEP high Eo high vs all other patients (AUC 50.6%, $p = 0.952$), **(D)** serum periostin protein as a predictor of CEP high Eo high vs CEP low Eo low groups (AUC 55.6%, $p = 0.614$), **(E)** sputum CLCA1 mRNA expression as a predictor of CEP high Eo high vs all other patients (AUC = 74.2%, $p = 0.01$) and **(F)** CEP high Eo high vs CEP low Eo low groups (AUC = 81.4%, $p = 0.007$).