



# The sputum transcriptome better predicts COPD exacerbations after the withdrawal of inhaled corticosteroids than sputum eosinophils

Benedikt Ditz<sup>1,2</sup>, Aartik Sarma<sup>3</sup>, Huib A.M. Kerstjens<sup>1,2</sup>, Jeroen J.W. Liesker<sup>1,2</sup>, Erik Bathoorn<sup>4</sup>, Judith M. Vonk<sup>1,2,5</sup>, Victor Bernal<sup>6</sup>, Peter Horvatovich<sup>6</sup>, Victor Guryev<sup>7</sup>, Saharai Caldera<sup>8,9</sup>, Chaz Langelier<sup>8,9</sup>, Alen Faiz<sup>1,2,10,11</sup>, Stephanie A. Christenson<sup>3,11</sup> and Maarten van den Berge<sup>1,2,11</sup>

<sup>1</sup>Dept of Pulmonary Diseases, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. <sup>2</sup>University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, The Netherlands. <sup>3</sup>University of California, San Francisco, CA, USA. <sup>4</sup>Dept of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. <sup>5</sup>Dept of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. <sup>6</sup>Dept of Analytical Biochemistry, Groningen Research Institute of Pharmacy, Groningen, The Netherlands. <sup>7</sup>European Research Institute for the Biology of Ageing, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. <sup>8</sup>Division of Infectious Diseases, Dept of Medicine, University of California, San Francisco, CA, USA. <sup>9</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA. <sup>10</sup>University of Technology Sydney, Respiratory Bioinformatics and Molecular Biology (RBMB), School of Life Sciences, Sydney, Australia. <sup>11</sup>These authors contributed equally.

Corresponding author: Maarten van den Berge (E-mail: [m.van.den.berge@umcg.nl](mailto:m.van.den.berge@umcg.nl))



Shareable abstract (@ERSpublications)

Sputum gene expression may have utility in biomarker development for identifying subjects who are at higher risk of exacerbation after ICS withdrawal <https://bit.ly/3gYl2OL>

Cite this article as: Ditz B, Sarma A, Kerstjens HAM, *et al.* The sputum transcriptome better predicts COPD exacerbations after the withdrawal of inhaled corticosteroids than sputum eosinophils. *ERJ Open Res* 2021; 7: 00097-2021 [DOI: 10.1183/23120541.00097-2021].

Copyright ©The authors 2021

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact [permissions@ersnet.org](mailto:permissions@ersnet.org)

This article has supplementary material available from [openres.ersjournals.com](https://openres.ersjournals.com)

Received: 8 Feb 2021  
Accepted: 26 April 2021

## Abstract

**Introduction** Continuing inhaled corticosteroid (ICS) use does not benefit all patients with COPD, yet it is difficult to determine which patients may safely sustain ICS withdrawal. Although eosinophil levels can facilitate this decision, better biomarkers could improve personalised treatment decisions.

**Methods** We performed transcriptional profiling of sputum to explore the molecular biology and compared the predictive value of an unbiased gene signature *versus* sputum eosinophils for exacerbations after ICS withdrawal in COPD patients. RNA-sequencing data of induced sputum samples from 43 COPD patients were associated with the time to exacerbation after ICS withdrawal. Expression profiles of differentially expressed genes were summarised to create gene signatures. In addition, we built a Bayesian network model to determine coregulatory networks related to the onset of COPD exacerbations after ICS withdrawal.

**Results** In multivariate analyses, we identified a gene signature (*LGALS12*, *ALOX15*, *CLC*, *IL1RL1*, *CD24*, *EMR4P*) associated with the time to first exacerbation after ICS withdrawal. The addition of this gene signature to a multiple Cox regression model explained more variance of time to exacerbations compared to a model using sputum eosinophils. The gene signature correlated with sputum eosinophil as well as macrophage cell counts. The Bayesian network model identified three coregulatory gene networks as well as sex to be related to an early *versus* late/nonexacerbation phenotype.

**Conclusion** We identified a sputum gene expression signature that exhibited a higher predictive value for predicting COPD exacerbations after ICS withdrawal than sputum eosinophilia. Future studies should investigate the utility of this signature, which might enhance personalised ICS treatment in COPD patients.

## Introduction

Inhaled corticosteroids (ICSs) are commonly used in the treatment of patients with COPD. However, only a subset of COPD patients benefits clinically from ICS, whereas they are ineffective and associated with adverse effects in others. Nevertheless, many patients use them and so far, it has been difficult to



determine which patients may safely sustain ICS withdrawal [1, 2]. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines recommends initiation of ICS treatment in combination with long-acting bronchodilators (LABAs) in COPD patients with a history of recurrent or severe exacerbations, a history of concomitant asthma, or blood eosinophil levels  $>300$  cells· $\mu\text{L}^{-1}$  [2]. Whether blood eosinophil counts are sufficient to guide ICS positioning in COPD patients is, however, debated [1–4]. Associations between blood eosinophils and clinical outcomes, including exacerbations, are not congruent across all observational cohort studies [1]. HASTIE *et al.* [3] have shown in the SPIROMICS cohort (N=827) that sputum eosinophils are poorly correlated with blood eosinophils but are a better predictor of COPD exacerbations.

Applying RNA-sequencing (RNA-seq) allows us to assess genome-wide expression profiles related to airway inflammation, which could be a useful tool to enhance personalised treatment decisions regarding ICS treatment in COPD. Whole-transcriptomic profiling in sputum represents a promising approach that might provide further insight into the molecular biology related to ICS treatment responses in COPD [5, 6]. Using this approach, BAINES *et al.* [6] identified a sputum gene signature that discriminated inflammatory phenotypes and predicted ICS treatment responses in asthma.

We previously showed in the SYMBEXCO trial that the likelihood of exacerbations after stopping ICS in COPD patients could be predicted by assessing sputum inflammation, particularly sputum eosinophil levels [7]. The SYMBEXCO trial was a randomised study comparing the efficacy of budesonide/formoterol *versus* prednisolone *versus* placebo during 2 weeks in COPD patients once they had an acute exacerbation [8]. In the current manuscript, we applied sputum transcriptome profiling (RNA-seq) in the same cohort to explore molecular networks related to exacerbations after ICS withdrawal. Further, we investigated whether an unbiased gene signature could be identified to further improve the prediction of exacerbations after ICS withdrawal, compared to sputum eosinophils. In addition, we performed Bayesian network modelling to explore coregulatory networks related to the onset of COPD exacerbations after ICS withdrawal.

## Methods

### Study design and sample collection

The study design of the SYMBEXCO trial (ClinicalTrials.gov Identifier: NCT00259779) has been described in detail elsewhere [7]. Briefly, COPD subjects were enrolled in a prospective cohort and followed for multiple consecutive visits until they developed an exacerbation and were randomised to treatment with prednisolone, budesonide/formoterol, or placebo [8]. When subjects were on ICS treatment at the time of enrolment, this was discontinued after their initial study visit. This manuscript presents the analysis of the period from the baseline visit to the time of the first exacerbation of those subjects in whom ICS treatment was withdrawn and who were able to produce an adequate sputum sample at baseline. Subjects who did not experience an exacerbation were followed until the end of their study participation. All participants were current or former smokers without a history of asthma or other significant respiratory diseases. The use of oral corticosteroids was not allowed for at least 4 weeks before inclusion. At each study visit, subjects underwent lung function testing followed by sputum induction with nebulised 4.5% sodium chloride solution for 3×5 min, with an adapted protocol for safety reasons when forced expiratory volume in 1 s (FEV<sub>1</sub>) was below 1.5 L [9]. An exacerbation was defined by increased breathlessness and at least two of the following symptoms for 24 h or more: increased cough frequency or severity, increased sputum volume or purulence, and increased wheeze, requiring extra prednisolone and/or antibiotics after ICS discontinuation as judged by a medical doctor [10]. The local medical ethics committee approved the study.

### Sample processing and sequencing

RNA was isolated using a RNeasy Mini kit (QIAGEN, Venlo, the Netherlands) and cDNA was synthesised and stored at  $-80^{\circ}\text{C}$  as described before [8, 11]. A NEBNext Ultra II Library Prep Kit (New England Biolabs, Ipswich, MA, USA) was used to prepare RNA-seq libraries, including fragmentation, end repair, adaptor ligation, and barcoding. DASH was employed to deplete unwanted rRNA sequences. Pre-amplification library preparation was completed in a single batch using an Echo liquid handling robot (Labcyte). RNA libraries were sequenced on a NovaSeq sequencer (Illumina, San Diego, CA, USA) by 150-bp paired-end sequencing. As a quality control step, every sputum sample was sequenced twice. Quality control was conducted on the raw sequence data using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and RNASeqC [12]. The Spliced Transcripts Alignment to a Reference (STAR) version 2.4.2a was used to align and identify all reads that belong to the human genome [13]. Samples were excluded if they had fewer than 10 000 transcripts with a minimum of eight copies.

### Data analysis

#### Discovery of the gene signature

Data were analysed with R statistical software version 3.6.1 (figure S1). Sputum RNA-seq data were analysed for the association between gene expression level and time to exacerbation after ICS withdrawal, using the likelihood ratio testing method in the edgeR package (R-package version 3.26.6) [14]. The variable 'time to exacerbation' (days) was  $\log_2$ -transformed to normalise the distribution and contained the time to first exacerbation or the time of monitoring for participants who did not experience an exacerbation before the end of their study participation (figure S2). We adjusted for smoking status, age, and sex and maintained a genome-wide false discovery rate (FDR) below 0.1 to control for multiple testing [15]. Genes significantly associated with time to exacerbation were assigned to signatures of relatively upregulated and downregulated genes. These signatures were separately summarised by using gene set variation analysis (GSVA) [16], and a gene set enrichment score was calculated per subject. The obtained gene signature enrichment scores (ESs) were compared to each other, time to exacerbation, absolute counts ( $\log_{10}$  transformed) of sputum inflammatory cells (eosinophils, lymphocytes, macrophages and neutrophils), and supernatant proteins (eosinophilic cationic protein and leukotriene-B4), using Spearman correlation testing.

#### Survival analyses

Next, univariate Cox regression analyses were performed to determine the hazards of experiencing an exacerbation after ICS withdrawal, for the ES of a derived gene signature (stratified into tertiles), sputum eosinophil percentages (stratified into tertiles), sex, smoking status, and history of exacerbations. Further, the clinical covariates pack-years of smoking (dichotomised by  $\geq$  or  $<$  median), sputum eosinophil percentages ( $\geq$  or  $<$ 3%), and the season of ICS withdrawal (dichotomised by November, December, January as opposed to outside those months) were investigated, which were previously identified as significant hazards in the SYMBEXCO trial [7]. Subsequently, significant hazards ( $p < 0.05$ ) were included in a multiple Cox regression model to adjust for potential confounders.

#### Bayesian network modelling

To determine coregulatory networks related to the onset of COPD exacerbations after ICS withdrawal, we integrated the normalised (FPKM method) sputum RNA-seq data in a Bayesian network model. The variable "time to exacerbation" was dichotomised ( $>$  or  $<$  mean) into an early ( $N=20$ ) and late/nonexacerbation phenotype ( $n=23$ ). We pre-selected the top 600 genes that were found significant (nominal  $p < 0.05$ ) in our likelihood ratio analysis. Their expression profiles were used as input together with the exacerbation phenotype as well as the demographic covariates age, sex, and smoking status. We built a Bayesian network using CGBayesnet version 7.14.14 in MATLAB version R2017b with the calculation of the exacerbation phenotype as binary primary phenotype and using the exhaustive search option [17].

#### Evaluation of the predictive value of the obtained gene signature for ICS-induced improvement in lung function in an independent RNA-seq dataset

The predictive value of the obtained gene signature was assessed in an independent RNA-seq dataset including airway epithelial brushings from 16 subjects with stable asthma, before and after treatment with inhaled budesonide, 180  $\mu\text{g}$  twice daily for 8 weeks. Additional details on this study are reported elsewhere [18]. The expression of the signature genes was summarised by using GSVA and the obtained gene signature ESs from samples before treatment with ICS were compared to improvement in  $\text{FEV}_1$  after treatment, using Spearman correlation testing. Also, the ESs of the gene signatures were compared to each other as well as to blood eosinophil levels. Further, the ICS sensitivity of the expression of the signature genes was assessed, both in SYMBEXCO (before and after 8 weeks of ICS withdrawal) as well as in the independent RNA-seq dataset (before and after 8 weeks of ICS treatment).

### Results

RNA-seq expression data from induced sputum samples at baseline were available for 45 participants. Two subjects restarted ICS treatment during study participation on request of their general physician, but without meeting the criteria of an exacerbation and were excluded from further analysis. Table 1 summarises the baseline characteristics of the 43 included participants, who were monitored for a median time of 144 days (interquartile range: 71–244). 38 participants experienced an exacerbation after a median time of 120 days (interquartile range: 69–203).

#### Association of baseline gene expression with time to exacerbation after ICS withdrawal

In total, 16094 genes were present with sufficiently large counts (filtering according to the default method in edgeR) to be analysed in association with time to exacerbation after ICS withdrawal. We identified six genes (*LGALS12*, *ALOX15*, *CLC*, *IL1RL1*, *CD24*, *EMR4P*) of which a higher expression prior to ICS

TABLE 1 Baseline participant clinical and demographic characteristics

Number of patients	43
Current/former smokers	23/20
Number of exacerbations	38
Males/females	36/7
Number of exacerbations in the past 12 months)	
0	19
1	17
$\geq 2$	7
Age (years)	64.5 (59.4–70.8)
Smoking history (pack-years)	38.0 (25.5–53.0)
Body mass index ( $\text{kg}\cdot\text{m}^{-2}$ )	25.9 (23.6–28.0)
Daily dose of ICS ( $\mu\text{g}$ ) <sup>#</sup>	800 (500–1000)
Pre-study treatment regimen	
ICS <sup>¶</sup>	8
ICS+LABA <sup>¶</sup>	30
ICS+LAMA <sup>¶</sup>	2
ICS+LAMA+LABA <sup>¶</sup>	3
Post-bronchodilator lung function	
FEV <sub>1</sub> (% pred)	64.1 (52.0–76.1)
FEV <sub>1</sub> /FVC	0.52 (0.44–0.56)
TLC (% pred)	112.1 (105.9–122.0)
RV (% pred)	135.5 (124.9–158.5)
RV/TLC (% pred)	114.9 (103.4–127.4)
Baseline inflammatory characteristics	
Sputum cells	
Eosinophils (%)	1.8 (0.7–4.0)
Neutrophils (%)	72.8 (67.0–80.3)
Macrophages (%)	18.0 (13.4–25.3)
Lymphocytes (%)	0.7 (0.1–1.2)
Bronchial epithelial cells (%)	2.0 (0.65–3.55)
Sputum supernatant proteins	
ECP ( $\mu\text{g}\cdot\text{L}^{-1}$ )	194.0 (58.6–718.0)
LTB4 ( $\text{ng}\cdot\text{mL}^{-1}$ )	0.49 (0.28–1.21)

Data are presented as n or median (interquartile range). ICS: inhaled corticosteroid; LABA: long-acting  $\beta$ -agonist; LAMA: long-acting muscarinic antagonist; FEV<sub>1</sub>: forced expiratory volume in 1 s; FVC: forced vital capacity; TLC: total lung capacity; RV: residual volume; ECP: eosinophilic cationic protein; LTB4: leukotriene-B4. #: daily dose of ICS calculated for indexed budesonide equivalent; ¶: additional short-acting  $\beta$ -agonist and/or short-acting muscarinic antagonist as needed.

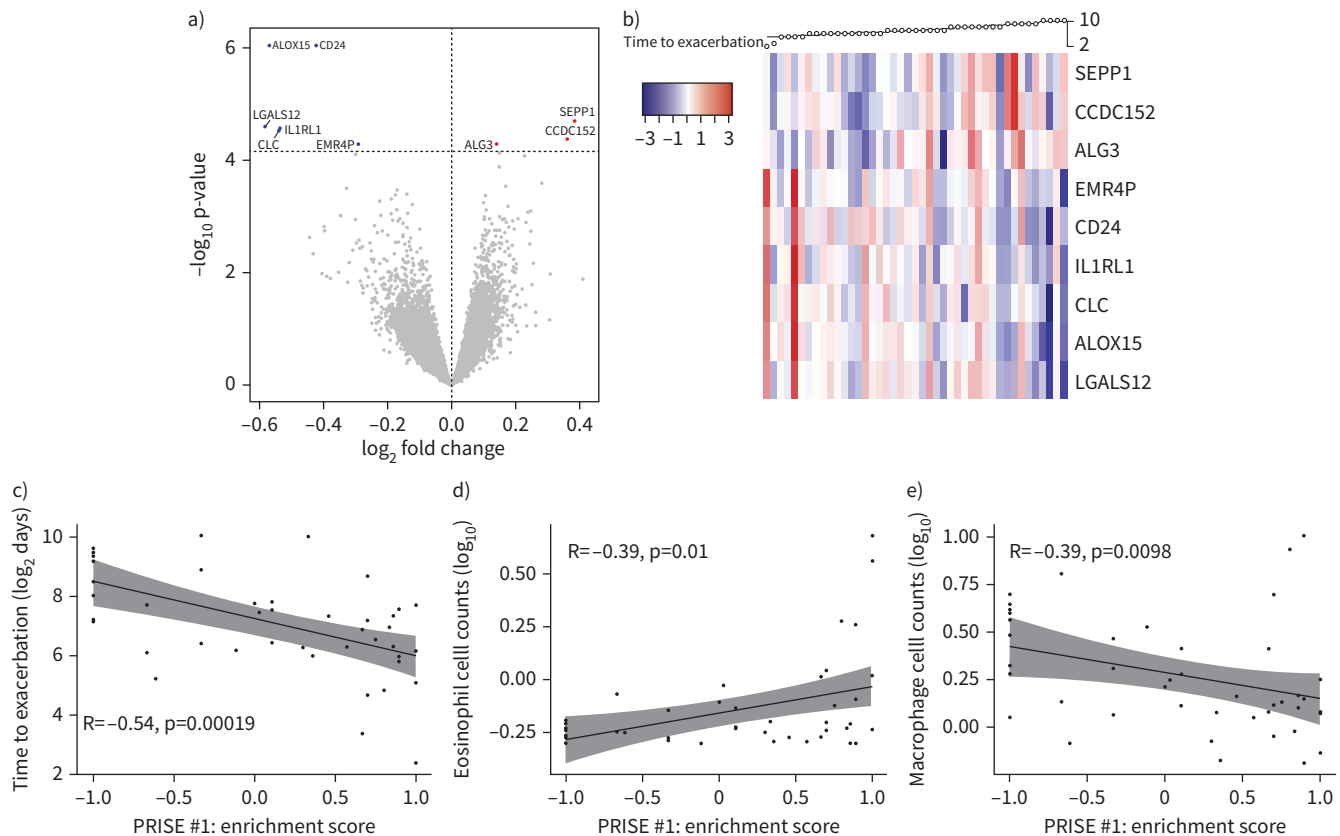
withdrawal was associated with a relatively short time to exacerbation, and this set of genes was then further analysed as *PRISE #1* (coPd exaceRbation Ics-withdrawal Sputum gEne signature) (genome-wide FDR<0.1) (figure 1 and table S1). In addition, we identified three genes (*SEPP1*, *CCDC152*, *ALG3*) of which a higher baseline expression was associated with a relatively long time to exacerbation, and this set of genes was then further analysed as *PRISE #2* (FDR<0.1) (figure 1 and table S1). Subsequently, ESs were calculated for each gene signature.

#### *PRISE #1 and #2 were highly correlated with each other as with time to exacerbation, sputum eosinophil and macrophage cell counts*

The ES of *PRISE #1* was highly correlated with time to exacerbation ( $\rho = -0.54$ ;  $p = 0.00019$ ) as well as sputum eosinophil ( $\rho = 0.39$ ;  $p = 0.01$ ), and macrophage cell counts ( $\rho = -0.39$ ;  $p = 0.0098$ ) (figure 1 and table S2). Further, the ES of *PRISE #2* was highly correlated with *PRISE #1* ( $\rho = -1.0$ ;  $p < 2.2 \times 10^{-16}$ ) as well as with sputum eosinophil ( $\rho = -0.37$ ;  $p = 0.014$ ), and macrophage cell counts ( $\rho = 0.4$ ;  $p = 0.0071$ ) (table S2). Both gene signatures did not correlate with lymphocyte and neutrophil cell counts or eosinophilic cationic protein and leukotriene-B4 levels (table S2).

#### *Monivariate Cox regression analyses*

Since both gene signatures were so highly correlated, the hazards of experiencing an exacerbation after ICS withdrawal were only determined for *PRISE #1*. Cox regression results, regarding *PRISE #2*, can be found



**FIGURE 1** Differential gene expression of sputum RNA-sequencing data. **a)** Volcano-plot of differentially expressed genes concerning time to exacerbation after inhaled corticosteroid (ICS) withdrawal (false discovery rate  $< 0.1$ ). Blue and red dots represent differentially expressed genes with negative and positive  $\log_2$  fold change, respectively. **b)** Heat map of differentially expressed genes. Subjects were clustered according to time to exacerbation (days;  $\log_2$  transformed). **c–e)** Spearman correlation testing comparing the enrichment score of *PRISE* #1 with the **c)** time to exacerbation, **d)** sputum eosinophil counts ( $\log_{10}$ ), and **e)** macrophage cell counts ( $\log_{10}$ ).

in the online supplementary material. The ESs of *PRISE* #1 were divided into three equal tertiles and tested in a univariate Cox regression model, where the first tertile served as a reference which was compared to the second and third tertile. The patients in the third tertile ( $ES \geq 0.69$ ) had a higher risk for developing a COPD exacerbation after ICS withdrawal compared with patients in the first tertile (hazard ratio (HR) 3.7,  $p = 0.003$ , the proportion of the variance explained ( $R^2$ ) = 0.202; table 2 and figure 2). Also, the covariates sex (male: HR 0.31,  $p = 0.012$ ,  $R^2 = 0.12$ ), history of exacerbations (HR 1.5,  $p = 0.046$ ,  $R^2 = 0.08$ ), season of ICS withdrawal (HR 0.3,  $p = 0.004$ ,  $R^2 = 0.152$ ), and pack-years of smoking  $\geq 38$  (HR 0.49,  $p = 0.038$ ,  $R^2 = 0.095$ ) were significant predictors, whereas smoking status was not (current smoking: HR 0.68,  $p = 0.024$ ,  $R^2 = 0.03$ ). Further, sputum eosinophil percentages were tested with two independent cut-offs. Sputum eosinophils  $\geq 3\%$  at the time of ICS withdrawal was associated with a significantly increased hazard (HR 2.3,  $p = 0.021$ ,  $R^2 = 0.11$ ) of experiencing a COPD exacerbation after ICS withdrawal, representing the cut-off that was previously identified in the SYMBEXCO trial [7]. In addition, sputum eosinophils percentages were divided into three equal tertiles, and patients in the third tertile had a higher risk for developing a COPD exacerbation after ICS withdrawal compared to patients in the first tertile (HR 2.8,  $p = 0.02$ ,  $R^2 = 0.124$ ).

#### Multiple Cox regression analyses

Next, we performed a multiple Cox regression analysis, including *PRISE* #1, sputum eosinophil percentages (dichotomised by  $<$  or  $\geq 3\%$ ) as well as the covariates sex, history of exacerbations, the season of ICS withdrawal, and pack-years of smoking. In this model, *PRISE* #1, but not sputum eosinophil level, remained statistically significant, as did sex, history of exacerbation, and the season of ICS withdrawal (table S3). When sputum eosinophil percentages were stratified by tertiles, instead of dichotomised by 3% similar results were obtained. Further, we determined the difference in explained variance (generalised  $R^2$ )

TABLE 2 Monovariate Cox regression hazard ratios

	Hazard ratio (95% CI)	p-value	Generalised R <sup>2</sup>
History of exacerbations	1.5 (1–2.1)	0.046*	0.08
Season of ICS withdrawal (not in November, December or January <sup>#</sup> )	0.3 (0.13–0.68)	0.004*	0.152
Sex (male)	0.31 (0.13–0.78)	0.012*	0.12
Smoking status (current smoker)	0.68 (0.36–1.30)	0.24	0.03
Pack-years smoking $\geq 38$ <sup>¶</sup>	0.49 (0.25–0.96)	0.038*	0.095
Sputum eosinophils $\geq 3\%$ <sup>+</sup>	2.3 (1.1–4.6)	0.021*	0.11
% Sputum eosinophils (stratified by tertiles)			
Second tertile (N=15)	1.4 (0.62–3.2)	0.42	0.124
Third tertile (N=15)	2.8 (1.17–6.5)	0.02*	
<b>PRISE #1 (ES stratified by tertiles)</b>			
Second tertile <sup>§</sup> (N=13)	2.3 (0.99–5.2)	0.052	0.202
Third tertile <sup>§</sup> (N=15)	3.7 (1.57–8.5)	0.003*	

ICS: inhaled corticosteroid; ES: enrichment score. <sup>#</sup>: dichotomised as “outside” versus “in” the period; <sup>¶</sup>: dichotomised using the median value at baseline; <sup>+</sup>: dichotomised  $\geq 3\%$  (N=15),  $<3\%$  (N=28). <sup>§</sup>: compared to the first tertile. \*: p<0.05.

between *PRISE* #1, and sputum eosinophil percentages in two separate multiple linear models, including either sputum eosinophil percentages or *PRISE* #1 (table 3). The multiple linear model, including *PRISE* #1, explained a higher percentage of the variance for experiencing an exacerbation after ICS withdrawal than sputum eosinophils percentages, *i.e.* 49.5% versus 37.7%.

#### Bayesian network modelling

The gene network associated with the binary exacerbation phenotype consisted of 86 nodes and 112 edges (figure 3). The exacerbation phenotype was linked to three subnetworks directly connected through *IL1RL1*, *SPRR2E*, and *TATDN3*, respectively. Sex was the only demographic covariate related to binary exacerbation phenotype that was identified in the network analysis, next to gene expression profiles. Here, six out of seven female subjects belonged to the early exacerbation phenotype (figure S5).

#### Evaluating *PRISE* #1 and #2 in an independent RNA-seq dataset

The predictive value of *PRISE* #1 and #2 was assessed in an independent RNA-seq dataset. There was no information on the expression of *CD24* as part of *PRISE* #1 in this dataset. The ESs of *PRISE* #1 and #2 from samples before ICS treatment were not associated with a change in FEV<sub>1</sub> (*PRISE* #1: rho=−0.28; p=0.47; *PRISE* #2: rho=0.28; p=0.47) after ICS treatment for 8 weeks (figure S6). However, the strong

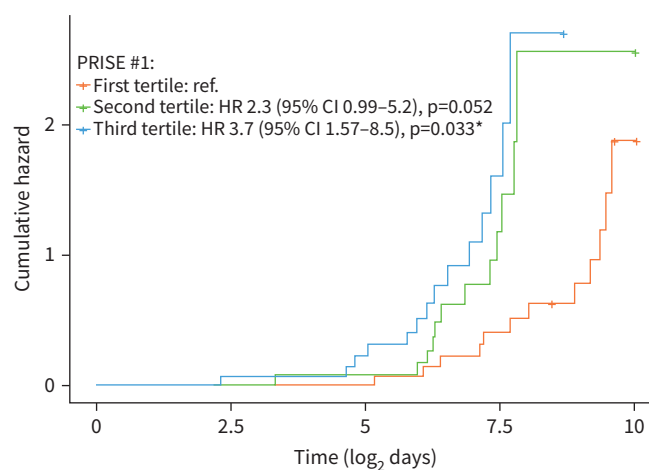


FIGURE 2 Cumulative hazard plots of risk of exacerbations. *PRISE* #1 enrichment scores are stratified by tertiles. The first tertile served as the reference for second and third tertiles. HR: hazard ratio. \*: p<0.05.





TABLE 4 ICS sensitivity of PRISE #1 and #2 genes in SYMBEXCO and independent RNA-sequencing dataset

Gene	SYMBEXCO: ICS treatment → no ICS (8 weeks)		Independent RNA-seq dataset: no ICS → ICS treatment (8 weeks)	
	Log <sub>2</sub> fold change	Nominal p-value	Log <sub>2</sub> fold change	Nominal p-value
CLC (PRISE #1)	0.76	<0.01*	−2.88	0.01*
LGALS12 (PRISE #1)	0.52	0.01*	−1.52	0.01*
CD24 (PRISE #1)	0.30	0.04*	NA	NA
ALOX15 (PRISE #1)	0.42	0.04*	−1.17	<0.01*
EMR4P (PRISE #1)	0.22	0.13	−0.63	0.16
IL1RL1 (PRISE #1)	0.15	0.36	−1.32	0.03*
ALG3 (PRISE #2)	−0.02	0.85	−0.49	0.02*
CCDC152 (PRISE #2)	−0.02	0.91	0.27	0.16
SEPP1 (PRISE #2)	0.01	0.96	−0.22	0.08

ICS: inhaled corticosteroid. \*: p<0.05.

after 8 weeks of ICS treatment). In SYMBEXCO, transcriptional profiles of four out six PRISE #1 genes increased after 8 weeks of ICS withdrawal (*CLC*, *LGALS12*, *CD24*, *ALOX15*) (table 4). *EMR4P* and *IL1RL1* showed the same trend, but missed statistical significance. The transcriptional profiles of PRISE #2 genes did not change. In the independent RNA-seq dataset, transcriptional profiles of four out six PRISE #1 genes decreased after 8 weeks of ICS treatment (*CLC*, *LGALS12*, *ALOX15*, *IL1RL1*). *EMR4P* showed the same trend, but missed statistical significance. The change of transcriptional profiles of PRISE #2 genes was inconsistent, since the expression of *ALG3* was decreased, whereas *CCDC152* and *SEPP1* did not change. In addition, the ESs of PRISE #1 decreased significantly after 8 weeks of ICS treatment (p=0.0042), whereas PRISE #2 showed the opposite trend, but missed statistical significance (p=0.013) (figure S8).

## Discussion

We identified nine genes predictive of an early exacerbation after ICS withdrawal. The addition of an unbiased gene signature to a multiple Cox regression model explained more variance in time to exacerbations compared to a model using sputum eosinophils. These findings suggest that sputum gene expression may have utility in biomarker development for identifying subjects who are at higher risk of exacerbation after ICS withdrawal.

PRISE #1 exhibited a high baseline expression, which was associated with a relatively early time to exacerbation. This indicates that a higher expression of the included genes is associated with an early exacerbation risk after ICS withdrawal. Several genes included in this signature have been previously identified to be involved in airway type 2 (T2) eosinophilic inflammation (*IL1RL1*, *LGALS12*, *EMR4P*, and *CLC* [19–21]). Other pathways of interest represented in the signature include macrophage-related inflammation (*IL1RL1*, *ALOX15* [22]), CD4<sup>+</sup> (*CLC* [23]), and CD8<sup>+</sup> T-cell differentiation (*IL1RL1*, *CD24* [24]).

*IL1RL1* (interleukin (IL)-1 receptor-like 1) encodes the ST2 receptor for IL-33 cytokine signalling, a key mechanism in airway T2 inflammation [21, 25]. Importantly, the *IL1RL1*–IL-33 axis does not only reflect T2 inflammation but is also involved in other immune cells and signalling pathways, including regulatory T-cell and macrophage-associated pathways [26]. *ALOX15* is expressed in monocyte-derived macrophages and thereby involved in orchestrating the nonimmunogenic removal of apoptotic cells as well as in facilitating inflammation resolution [22, 27]. *CD24* is involved in diverse functions of the adaptive immune response, including B-cells, neutrophils, and the regulation of CD8<sup>+</sup> T-cell activation through HMGB1-mediated engagement of T-cell RAGE [24, 28]. The CLC protein is a major constituent of eosinophils and is also expressed by basophils and CD4<sup>+</sup> CD25<sup>+</sup> regulatory T-cells [23, 29]. In asthma patients, its expression was decreased in corticosteroid responders, indicating its potential to predict clinical responses to corticosteroids [30].

Our Bayesian network model identified *IL1RL1* to be one of the key hub genes, followed by the remaining genes of PRISE #1. In addition, PRISE #1 correlated positively and negatively with sputum eosinophil and macrophage cell counts, respectively. Therefore, it could be speculated that a cascade, consisting of both eosinophils and macrophages as innate effector cells, is associated with the propensity to exacerbate in COPD patients after ICS withdrawal. Further, it is tempting to speculate that anti-*IL1RL1*/IL-33 may



represent a treatment option for frequent exacerbators with COPD, yet this remains to be investigated in future studies.

Two additional gene subnetworks were identified in the Bayesian network model, which were directly connected to an early *versus* late/nonexacerbation phenotype. Genes identified in *PRISE* #2 were connected via a *TATDN3* subnetwork, which is a protein-coding gene related to deoxyribonuclease activity. The third gene subnetwork was directly connected through *SPRR2E*, which encodes the small proline-rich protein 2E, and is involved in epithelial cell-related processes. *SPRR2E* expression was previously found to be increased in primary bronchial epithelial cells from subjects with COPD compared to healthy controls, suggesting that this gene might be involved in airway epithelial inflammatory or remodelling processes in COPD [31]. As additional genes of the multigene *SPRR* family, as well as keratin genes, were connected to this subnetwork, our results indicate that the airway epithelial cell function might be associated with the exacerbation phenotype after ICS withdrawal.

Importantly, sex, but not smoking status was another significant demographic predictor for experiencing an exacerbation after ICS withdrawal, which was also identified as a direct subnetwork related to the exacerbation phenotype in the Bayesian model. Sex-based differences in COPD exacerbation frequency have been shown previously, with a higher rate of exacerbations among women compared to men [32]. Our network model indicates that female COPD patients are more prone to early exacerbations after ICS withdrawal as six out of seven female subjects belonged to the early exacerbation phenotype. Importantly, this observation is based on a relatively low number of participants as only seven out of 43 participants were women, and therefore requires further investigation.

The predictive value of *PRISE* #1 and #2 could not be validated in an independent RNA-seq dataset concerning ICS-induced improvement in lung function; however, this analysis was likely underpowered. In addition, the clinical outcome, as well as airway specimen, differed compared to the SYMBEXCO dataset. Still, the strong associations between the ESs of *PRISE* #1 and #2 as well as eosinophils counts in blood were validated in this dataset. Also, we showed that the expression of *PRISE* #1 genes was steroid-sensitive, both in SYMBEXCO and the independent RNA-seq dataset, indicating that the expression of these genes increases with ICS withdrawal and decreases with ICS treatment, in subjects with COPD and asthma.

There are some limitations related to our study. The identified gene signature should be validated in an independent exacerbation cohort, which would be crucial to investigate its utility as a biomarker. To our knowledge, such a replication cohort is currently not available. Further, our results are based on a COPD study population with predominantly moderate airflow obstruction and it is not clear to what extent our observations also apply to milder or more severe COPD.

In summary, we identified a sputum gene signature that exhibited a higher predictive value for predicting COPD exacerbations after ICS withdrawal than sputum eosinophilia. Future studies should investigate the utility of this signature, which might enhance personalised ICS treatment in COPD patients.

Author contributions: J.J.W. Liesker, E. Bathoorn and H.A.M. Kerstjens conducted the SYMBEXCO study. A. Sarma, S. Caldera, C. Langelier and S.A. Christenson were responsible for the preparation and RNA sequencing of the sputum samples. B. Ditz, A. Sarma, J.M. Vonk, V. Bernal, P. Horvatovich and V. Guryev performed the statistical analysis of the data. B. Ditz wrote the initial draft of the manuscript with additional content provided and critical revisions from all authors.

Conflict of interest: B. Ditz has nothing to disclose. A. Sarma reports grants from National Heart, Lung, and Blood Institute, during the conduct of the study. H.A.M. Kerstjens reports a grant from AstraZeneca as well as grants and fees for consultancy or advisory board participation from GlaxoSmithKline, Boehringer Ingelheim, and Novartis, and a grant from Chiesi, all outside of the submitted work and all paid to his institution. J.J.W. Liesker has nothing to disclose. E. Bathoorn has nothing to disclose. J.M. Vonk has nothing to disclose. V. Bernal has nothing to disclose. P. Horvatovich has nothing to disclose. V. Guryev has nothing to disclose. S. Caldera has nothing to disclose. C. Langelier has nothing to disclose. A. Faiz has nothing to disclose. S.A. Christenson reports consulting fees from AstraZeneca, GlaxoSmithKline, Amgen and Glenmark; personal fees for invited lectures from Sunovion and Genentech; and personal fees for writing for UpToDate, all outside the submitted work. M. van den Berge has nothing to disclose.

Support statement: The submitted work was co-financed by the Ministry of Economic Affairs and Climate Policy by means of the PPP. A. Faiz was supported by a junior Longfond grant (4.2.16.132JO). Funding information for this article has been deposited with the Crossref Funder Registry.

## References

- 1 Agusti A, Fabbri LM, Singh D, *et al.* Inhaled corticosteroids in COPD: friend or foe? *Eur Respir J* 2018; 52: 1801219.
- 2 GOLD. 2020 Global Strategy for the Diagnosis, Management and Prevention of COPD. Global Initiative for Chronic Obstructive Lung Disease (GOLD), 2020. Available from: <https://goldcopd.org/gold-reports/>
- 3 Hastie AT, Martinez FJ, Curtis JL, *et al.* Sputum or blood eosinophil association with clinical measures of COPD severity in the SPIROMICS cohort. *Lancet Respir Med* 2017; 5: 956–967.
- 4 Casanova C, Celli BR, De-Torres JP, *et al.* Prevalence of persistent blood eosinophilia: relation to outcomes in patients with COPD. *Eur Respir J* 2017; 50: 1701162.
- 5 Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* 2012; 18: 716–725.
- 6 Baines KJ, Simpson JL, Wood LG, *et al.* Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. *J Allergy Clin Immunol* 2014; 133: 997–1007.
- 7 Liesker JJW, Bathoorn E, Postma DS, *et al.* Sputum inflammation predicts exacerbations after cessation of inhaled corticosteroids in COPD. *Respir Med* 2011; 105: 1853–1860.
- 8 Bathoorn E, Liesker JJW, Postma DS, *et al.* Anti-inflammatory effects of combined budesonide/formoterol in COPD exacerbations. *COPD J Chronic Obstr Pulm Dis* 2008; 5: 282–290.
- 9 Bathoorn E, Liesker J, Postma D, *et al.* Safety of sputum induction during exacerbations of COPD. *Chest* 2007; 131: 432–438.
- 10 Davies L, Angus RM, Calverley PMA. Oral corticosteroids in patients admitted to hospital with exacerbations of chronic obstructive pulmonary disease: a prospective randomised controlled trial. *Lancet* 1999; 354: 456–460.
- 11 Szafranski W, Cukier A, Ramirez A, *et al.* Efficacy and safety of budesonide / formoterol in the management of chronic obstructive pulmonary disease. *Eur Respir J* 2003; 21: 74–81.
- 12 Deluca DS, Levin JZ, Sivachenko A, *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 2012; 28: 1530–1532.
- 13 Dobin A, Davis CA, Schlesinger F, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15–21.
- 14 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; 26: 139–140.
- 15 Benjamini Y, Drai D, Elmer G, *et al.* Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001; 125: 279–284.
- 16 Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013; 14: 7.
- 17 McGeachie MJ, Chang H, Weiss ST. CGBayesNets: conditional Gaussian Bayesian network learning and inference with mixed discrete and continuous data. *PLoS Comput Biol* 2014; 10: e1003676.
- 18 Bhakta NR, Christenson SA, Nerella S, *et al.* IFN-stimulated gene expression, type 2 inflammation, and endoplasmic reticulum stress in asthma. *Am J Respir Crit Care Med* 2018; 197: 313–324.
- 19 Virkud YV, Kelly RS, Croteau-Chonka DC, *et al.* Novel eosinophilic gene expression networks associated with IgE in two distinct asthma populations. *Clin Exp Allergy* 2018; 48: 1654–1664.
- 20 Sridhar S, Liu H, Pham TH, *et al.* Modulation of blood inflammatory markers by benralizumab in patients with eosinophilic airway diseases. *Respir Res* 2019; 20: 1–12.
- 21 Rossios C, Pavlidis S, Hoda U, *et al.* Sputum transcriptomics reveal upregulation of IL-1 receptor family members in patients with severe asthma. *J Allergy Clin Immunol* 2018; 141: 560–570.
- 22 Snodgrass RG, Brüne B. Regulation and functions of 15-lipoxygenases in human macrophages. *Front Pharmacol* 2019; 10: 719.
- 23 Kubach J, Lutter P, Bopp T, *et al.* Human CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells: proteome analysis identifies galectin-10 as a novel marker essential for their anergy and suppressive function. *Blood* 2007; 110: 1550–1558.
- 24 Kim TS, Gorski SA, Hahn S, *et al.* Distinct dendritic cell subsets dictate the fate decision between effector and memory CD8<sup>+</sup> T cell differentiation by a CD24-dependent mechanism. *Immunity* 2014; 40: 400–413.
- 25 Gordon ED, Palandra J, Wesolowska-Andersen A, *et al.* IL1RL1 asthma risk variants regulate airway type 2 inflammation. *JCI Insight* 2016; 1: e87871.
- 26 Griesenauer B, Paczesny S. The ST2/IL-33 axis in immune cells during inflammatory diseases. *Front Immunol* 2017; 8: 475.
- 27 Kuhn H, Gehring T, Schröter A, *et al.* Cytokine-dependent expression regulation of ALOX15. *J Cytokine Biol Kuhn* 2016; 1: 106.
- 28 Fang X, Zheng P, Tang J, *et al.* CD24: from A to Z. *Cell Mol Immunol* 2010; 7: 100–103.
- 29 Ackerman BSJ, Weil GJ, Gleich GJ. Formation of Charcot-Leyden crystals by human basophils. *J Exp Med* 1982; 155: 1597–1609.

- 30 Berthon BS, Gibson PG, Wood LG, *et al.* A sputum gene expression signature predicts oral corticosteroid response in asthma. *Eur Respir J* 2017; 49: 1700180.
- 31 Baines KJ, Hsu AC, Tooze M, *et al.* Novel immune genes associated with excessive inflammatory and antiviral responses to rhinovirus in COPD. *Respir Res* 2013; 14: 15.
- 32 Hurst JR, Vestbo J, Anzueto A, *et al.* Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010; 363: 1128–1138.