Supplementary material

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

B. Ditz[1,2], A. Sarma[3], H.A.M. Kerstjens[1,2], J.J.W. Liesker[1,2], E. Bathoorn[4], J.M. Vonk[2,5], V. Bernal[6], P. Horvatovich[6], V. Guryev[7], S. Caldera[8,9], C. Langelier[8,9], A. Faiz[1,2,10*], S. A. Christenson[3*], M. van den Berge[1,2*]

METHODS

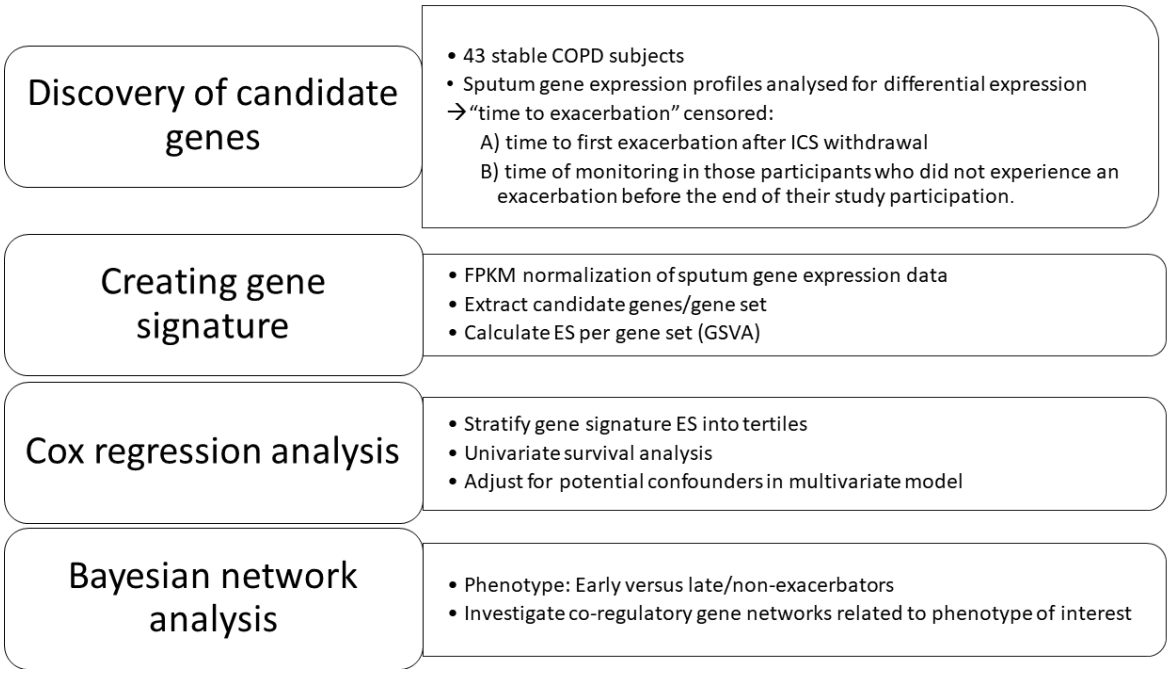

Figure S1. Schematic diagram representing the analyses of the RNA-seq expression data in this study.

RESULTS

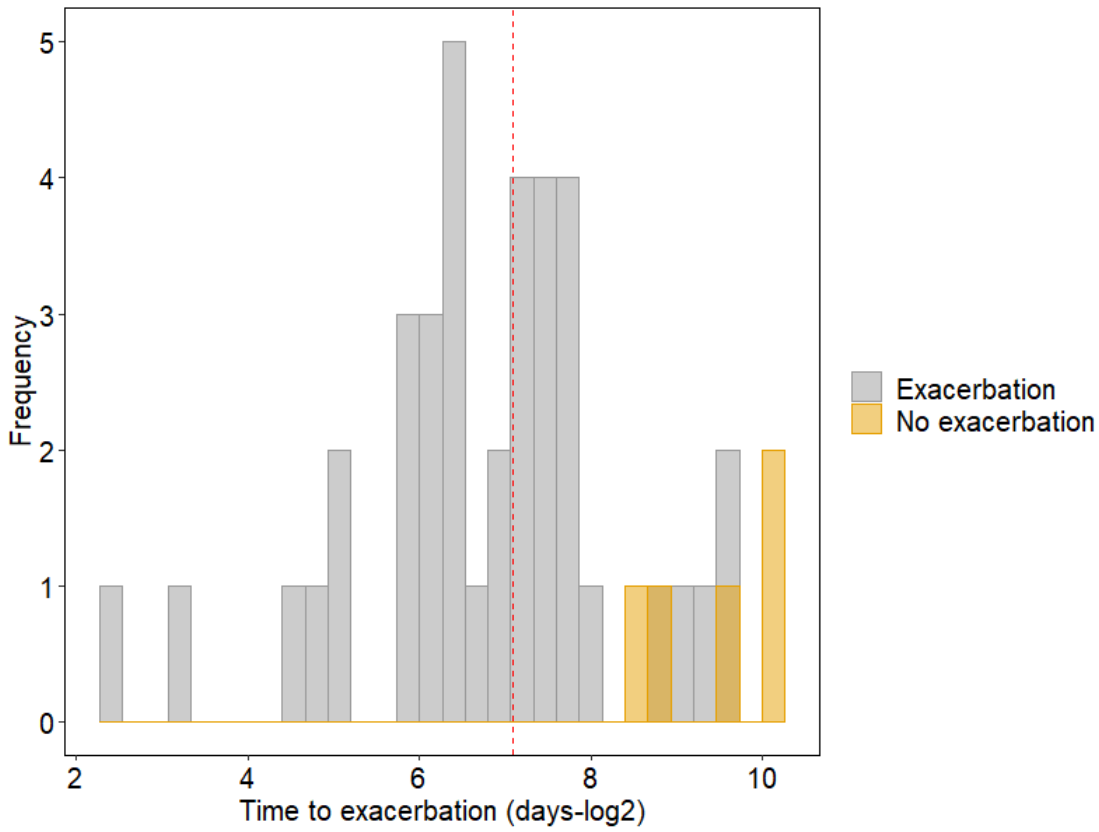

Figure S2. Histogram of the time to exacerbation. The time to exacerbation is censored containing the days to first exacerbation as well as monitoring time of participants not experiencing an exacerbation before the end of the study. Days of time to exacerbation are log2-transformed. Red dashed line represents mean. Subjects to its left belong to the early exacerbation phenotype (N=20), and subjects to its right belong to the late/non-exacerbation phenotype (N=23), which were integrated in the Bayesian network model.

Cox regression analyses including *PRISE* #2

The enrichment scores of *PRISE* #2 were divided into three equal tertiles and tested in an univariate cox regression model, where the first tertile served as the reference, which was compared to the second and third tertile. Patients in the third tertile (ES $\geq$ 0.45) had a lower risk for developing a COPD exacerbation after ICS withdrawal compared to the patients in the first tertile (HR 0.25, p=0.001, R2=0.218) (Figure S3). Next, we performed a multivariate Cox regression analysis, including *PRISE* #2, sputum eosinophil percentages (dichotomized by $</\geq$ 3%) as well as the clinical co-variates history of exacerbations,the season of ICS withdrawal, packyears of smoking. In this model, *PRISE* #2, but not sputum eosinophil percentages remained statistically significant, next to the history of exacerbation and the

season of ICS withdrawal (Table S4). Importantly, sputum eosinophil percentages, stratified by tertiles instead of dichotomized by 3%, did not change the outcome.

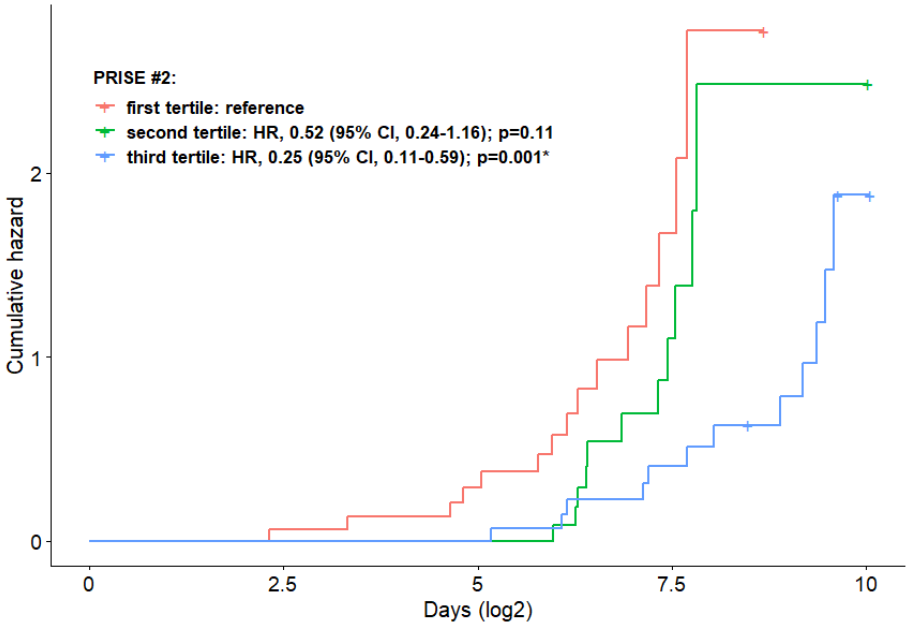

Figure S3. Cumulative hazard plots of risk of exacerbations. *PRISE* #2 enrichment scores (ES) are stratified by tertiles; HR=Hazard ratio; 95 % CI= 95% confidence interval; first tertile served as reference for second and third tertile, respectively; p=p-value; *= p-value significance (<0.05).

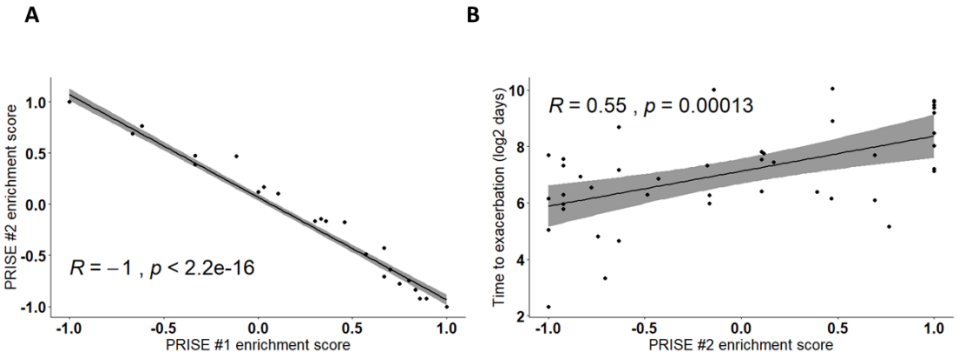

Figure S4. Correlations between both gene signature enrichment scores. Spearman correlation testing was applied, comparing the enrichment score (ES) of *PRISE* #1&2 as well as *PRISE* #2 and time to exacerbation

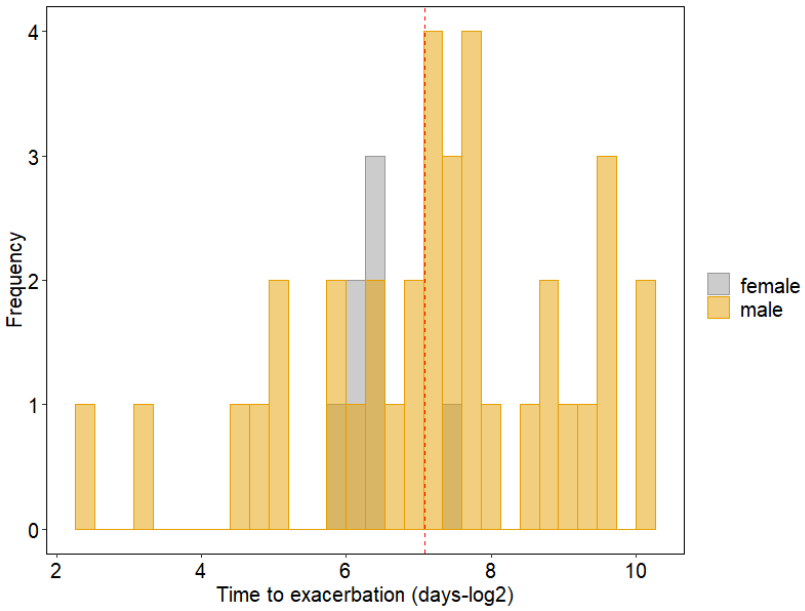

Figure S5. Histogram of sex distribution concerning the time to exacerbation. Time to exacerbation is censored containing the days to first exacerbation as well as monitoring time of participants not experiencing an exacerbation before the end of the study. Time to exacerbation is represented in days, which are log2-transformed. Red dashed line represents the mean.

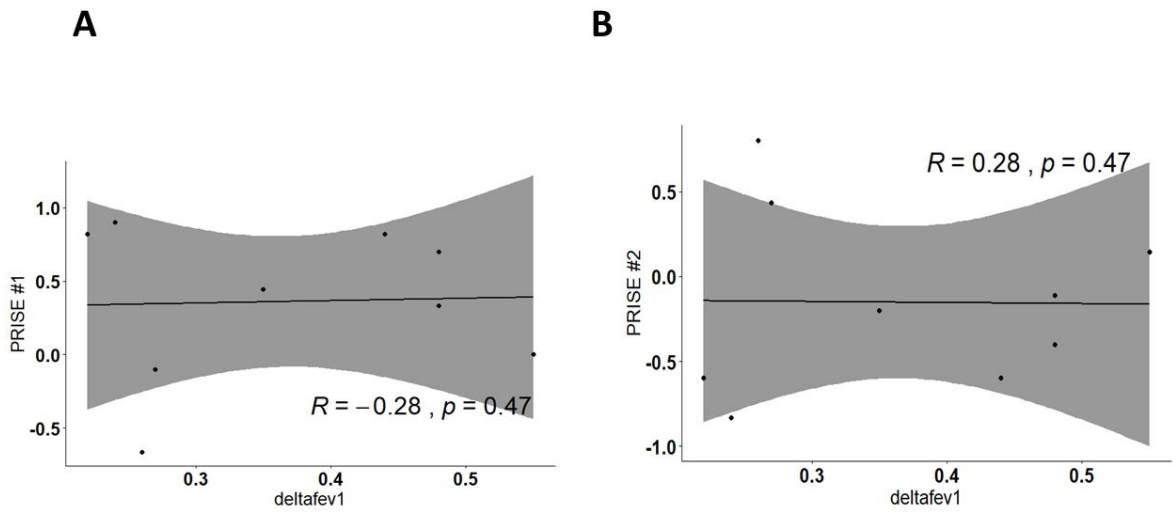

Figure S6. Correlations of PRISE #1&2 with ICS-induced improvement in FEV1 in independent RNA-seq dataset. Steroid responsiveness of PRISE #1 and #2 was investigated in an independent RNA-seq dataset with asthmatic patients, before and after treatment with inhaled budesonide, 180 mcg twice daily for 8 weeks. GSVA enrichment scores from samples before steroid treatment were compared to the improvement of in FEV1 in the independent dataset, using Spearman correlation testing.

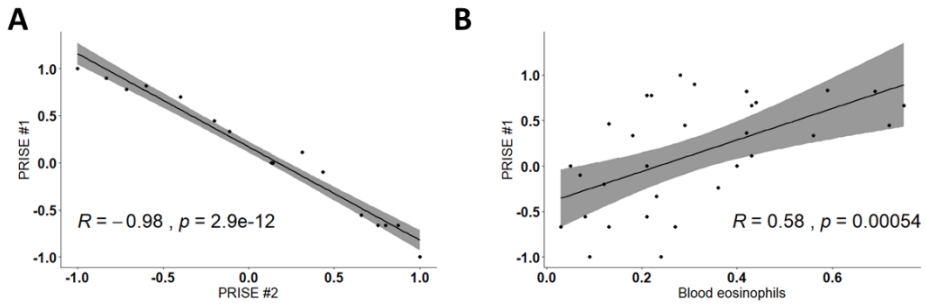

Figure S7. Correlations between PRISE #1&2 enrichment scores and blood eosinophils in independent RNA dataset. Spearman correlation testing was applied, comparing the enrichment score (ES) of *PRISE* #1&2 as well as *PRISE* #1 and blood eosinophil levels.

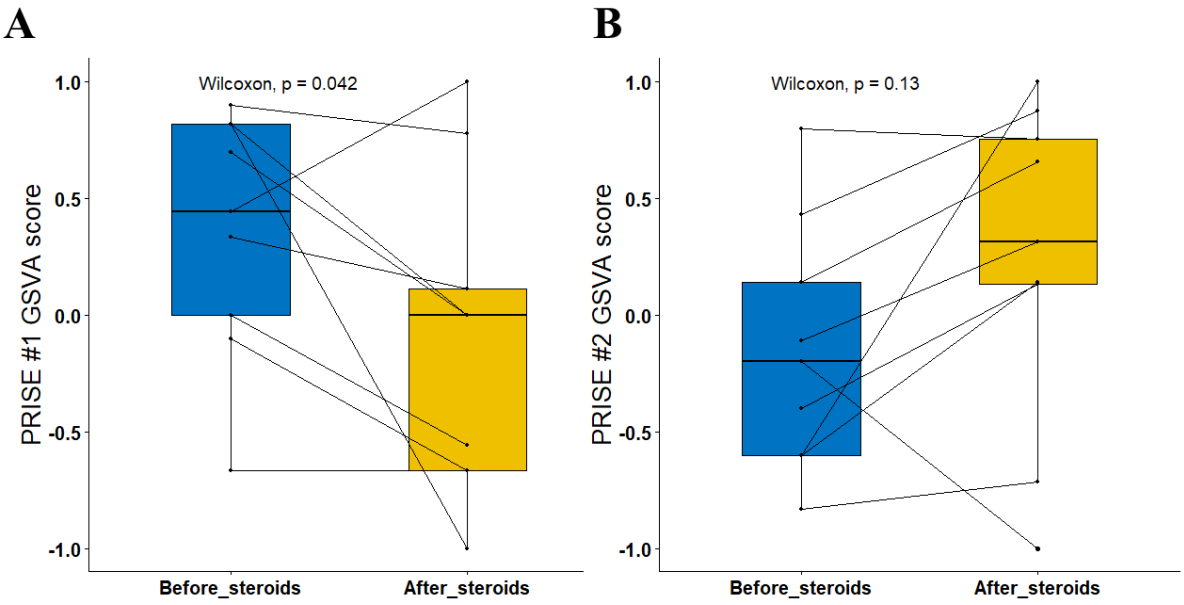

Figure S8. PRISE #1&2 enrichment scores before and after eight weeks ICS treatment in the independent RNA-seq dataset.

Enrichment scores of PRISE #1&2 were compared before and after eight weeks ICS treatment in theindependent RNA-seq dataset, using Wilcoxon signed ranked test.

Table S1. Differential gene expression of sputum RNA-seq data concerning time to exacerbation (FDR<0.1)

	logFC	logCPM	LR	PValue	FDR	human_gene	Gene signature #
ENSG00000133317	-0,58	3,40	17,73	2,54E-05	0,080	LGALS12	1
ENSG00000161905	-0,57	4,54	24,12	9,04E-07	0,007	ALOX15	1

ENSG00000105205	-0,54	3,34	17,44	2,97E-05	0,080	CLC	1
ENSG00000115602	-0,54	4,94	17,57	2,77E-05	0,080	IL1RL1	1
ENSG00000272398	-0,42	3,73	24,15	8,92E-07	0,007	CD24	1
ENSG00000268758	-0,29	4,01	16,37	5,21E-05	0,093	EMR4P	1
ENSG00000214160	0,14	3,96	16,39	5,15E-05	0,093	ALG3	2
ENSG00000198865	0,36	3,01	16,79	4,17E-05	0,093	CCDC152	2
ENSG00000250722	0,39	3,49	18,16	2,03E-05	0,080	SEPP1	2

Table S2. Correlations between *PRISE* # 1&2, and other sputum inflammatory markers

***PRISE* #1**	**Rho**	**p**	***PRISE* #2**	**Rho**	**p**
PRISE #2	-1	2.2e-16*			
Eosinophil cell counts	0.39	0.01*	Eosinophil cell counts	-0.37	0.014*
Macrophage cell counts	-0.39	0.0098*	Macrophage cell counts	0.4	0.0071*
Lymphocyte cell counts	-0.029	0.85	Lymphocyte cell counts	0.038	0.81
Neutrophil cell counts	-0.24	0.12	Neutrophil cell counts	0.26	0.09
ECP (mcg/L)	0.29	0.057	ECP (mcg/L)	0.049	0.77
LTB4 (ng/L)	-0.045	0.79	LTB4 (ng/L)	-0.27	0.076

Sputum inflammatory makers were $\log_{10}$ transformed; ECP= eosinophilic cationic protein; LTB4= leukotriene-B4; Rho=Spearman's Rank correlation coefficient; *indicates p value significance (<0.05)

Table S3. Multivariate Cox regression model including sputum eosinophils percentage and *PRISE* #1

		Hazard ratio	**95% CI**	**p-value**	**Generalized R^2**
History of exacerbations		1.53	1.03-2.28	0.036*	0.495
Season of ICS withdrawal (not in November, December or January[a])		0.27	0.10-0.69	0.006*	
Packyears smoking ≥ 38[b]		0.51	0.22-1.18	0.114	
Sputum eosinophils ≥ 3%[c]		1.05	0.45-2.44	0.908	
Sex (male)		0.46	0.17-1.30	0.143	
PRISE #1 (ES stratified by	Second tertile (N=13)	2.53	0.99-6.49	0.053	
	Third tertile	5.18	1.77-15.17	0.003*	

tertiles)	(N=15)				

[a]= Dichotomised as "outside" versus "in" the period; [b]=dichotomised using the median value at baseline; [c]= dichotomized, ≥ 3% (N=15), <3% (N=28); ES= enrichment score; *indicates p value significance (<0.05)

Table S4. Multivariate Cox regression model including sputum eosinophil percentage and *PRISE* #2

		Hazard ratio	**95% CI**	**p-value**	**Generalized R^2**
History of exacerbations		1.42	0.95-2.13	0.09	0.482
Season of ICS withdrawal (not in November, December or January[a])		0.30	0.11-0.76	0.012*	
Packyears smoking ≥ 38[b]		0.39	0.18-0.87	0.021*	
Sputum eosinophils ≥ 3%[c]		1.14	0.49-2.70	0.758	
PRISE #2 (ES stratified by tertiles)	Second tertile (N=9)	0.35	0.14-0.87	0.024	
	Third tertile (N=15)	0.18	0.07-0.50	<0.001*	

[a]= Dichotomised as "outside" versus "in" the period; [b]=dichotomised using the median value at baseline; [c]= dichotomized, ≥ 3% (N=15), <3% (N=28); *indicates p value significance (<0.05)

REFERENCE

1. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*. 2013;14. doi:10.1186/1471-2105-14-7

2. Liesker JJW, Bathoorn E, Postma DS, Vonk JM, Timens W, Kerstjens HAM. Sputum inflammation predicts exacerbations after cessation of inhaled corticosteroids in COPD. *Respir Med*. 2011;105(12):1853-1860. doi:10.1016/j.rmed.2011.07.002

3. 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(6). doi:10.1371/journal.pcbi.1003676