Biomarkers of chronic airflow limitation and COPD identified by mass spectrometry

Magnus Molin[1], Anne Incamps[2], Manon Lemasson[2], Mats Andersson[1], Eleftheria Pertsinidou[1,3], Marieann Högman[4], Karin Lisspers[5], Björn Ställberg[5], Anders Sjölander[1], Andrei Malinovschi[6,7] and Christer Janson[4,7]

[1]Thermo Fisher Scientific, Uppsala, Sweden. [2]Thermo Fisher Scientific, Nimes, France. [3]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. [4]Department of Medical Sciences, Respiratory, Allergy and Sleep Research, Uppsala University, Uppsala, Sweden. [5]Department of Public Health and Caring Sciences, Family Medicine and Preventive Medicine, Uppsala University, Uppsala, Sweden. [6]Department of Medical Sciences, Clinical Physiology, Uppsala University, Sweden. [7]These authors contributed equally.

Corresponding author: Andrei Malinovschi (andrei.malinovschi@medsci.uu.se)

Shareable abstract (@ERSpublications)
This study identified peptides for histidine-rich glycoprotein (HRG), α_1-acid glycoprotein (AGP1) and α_1-antitrypsin that separate chronic airflow limitation/COPD subjects from matched controls, outperforming other approaches. HRG and AGP1 are novel findings. https://bit.ly/3QAnafJ

Cite this article as: Molin M, Incamps A, Lemasson M, *et al*. Biomarkers of chronic airflow limitation and COPD identified by mass spectrometry. *ERJ Open Res* 2024; 10: 00751-2023 [DOI: 10.1183/23120541.00751-2023].

Received: 9 Oct 2023
Accepted: 7 Nov 2023

Abstract

Rationale COPD affects 300 million people worldwide and is the third leading cause of death according to World Health Organization global health estimates. Early symptoms are subtle, and so COPD is often diagnosed at an advanced stage. Thus, there is an unmet need for biomarkers that can identify individuals at early stages of the disease before clinical symptoms have manifested. To date, few biomarkers are available for clinical diagnostic use in COPD.

Methods We evaluated a panel of serum biomarkers related to inflammation and infection for their ability to discriminate between 77 subjects with chronic airflow limitation (CAL) and 142 subjects with COPD, *versus* 150 healthy subjects (divided into two control groups that were matched with regards to age, gender and smoking to CAL and COPD). Healthy subjects and CAL were from Burden of Obstructive Lung Disease (BOLD), a population-based study. CAL was defined by post-bronchodilatory forced expiratory volume in 1 s/forced vital capacity ratio <0.7 in the BOLD population. COPD subjects were from Tools for Identifying Exacerbations (TIE), a COPD patient cohort. Quantification of 100 biomarker candidates was done by liquid chromatography-tandem mass spectrometry.

Results Several protein-derived peptides were upregulated in CAL, compared to controls; most notably peptides representing histidine-rich glycoprotein (HRG), α_1-acid glycoprotein (AGP1), α_1-antitrypsin (α1AT) and fibronectin. Out of these, HRG-, AGP1- and α1AT-specific peptides were also elevated in the COPD cohort.

Conclusion HRG, AGP1 and α1AT biomarkers distinguish subjects with CAL and COPD from healthy controls. HRG and AGP1 represent novel findings.

Introduction

COPD affects 300 million people worldwide and is the third major cause of death, responsible for ~6% of total deaths [1]. COPD is a group of respiratory diseases (including emphysema and chronic bronchitis) characterised by chronic airflow limitation (CAL) and breathing difficulties and is typically caused by smoking and exposure of the airways to fumes and dust particles [2]. COPD develops over time and with increasing age and thus relates to ageing, and possibly to an accelerated lung ageing process. It has been described as a disease characterised by organ-specific senescence driven by stressors such as smoking-induced DNA damage (reviewed by Childs *et al.* [3]) and senescent cells have been shown to drive the pulmonary changes seen in COPD [4]. Early nonpharmacological intervention, such as smoking

cessation, can prevent lung function impairment and disease progression [5]. However, once clinical symptoms have manifested, there is no cure or medication to resolve COPD and since early symptoms are subtle, COPD is often diagnosed at an advanced stage.

Diagnosis of COPD is considered in patients with dyspnoea, chronic cough, chronic sputum production and history of exposure to risk factors, such as smoking and exposure to fumes and dust particles. Diagnosis includes assessment by spirometry where the presence of a post-bronchodilator forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) ratio <0.7 defines CAL, which confirms a COPD diagnosis in subjects with risk factors and symptoms according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) [6]. In addition, there are different severity stages of airflow limitation in COPD based on FEV_1 as defined by GOLD [6]. Spirometry is the most readily available, reproducible and objective measurement of airflow limitation. However, spirometry has limitations as it reflects to a lesser extent obstruction in peripheral airways and emphysema. Thus, there is an unmet need for biomarkers that can identify individuals at pre-clinical stages of the disease.

Several biomarker candidates are being studied and some have been evaluated for clinical use in COPD, *e.g.* fibrinogen and C-reactive protein (CRP) [7], but to date no single biomarker has been found to be a clinically useful tool in diagnosing COPD, mainly because of large overlap between healthy and COPD groups. Thus, even though patients with more severe COPD as a group have, for example, higher plasma fibrinogen levels than those with mild-to-moderate disease, variation around group mean values is too high to be clinically useful for the individual patient [8]. Nonetheless, inflammation in COPD is considered to drive disease progression [9]. Several studies have observed an increase in the blood of certain inflammatory mediators in COPD [7], some of which are so-called acute-phase proteins [10]. However, profiling of cellular and inflammatory changes in the airway wall has yielded conflicting results on when, during the disease course, and in what part of the airways different inflammatory cells and mediators either increase or decrease during disease onset and progression [11].

Fibrinogen and CRP are both acute-phase proteins. Acute-phase proteins are produced by different cells and tissues, most notably by immune cells and the liver, in response to an infection, inflammation or trauma. There are two main categories of acute-phase proteins: positive acute-phase proteins and negative acute-phase proteins. Positive acute-phase proteins, such as CRP, serum amyloid A and fibrinogen, are increased in concentration during acute inflammation. Their levels rise rapidly in response to inflammation and can serve as biomarkers for assessing the severity and progression of inflammatory conditions. Negative acute-phase proteins, such as albumin and transferrin, are reduced in concentration during the acute-phase response. This reduction occurs because the production of these proteins is suppressed while the body focuses its resources on producing positive acute-phase proteins.

We speculated that the acute-phase response is triggered early during disease initiation and that components of the response could serve as early indicators of airway obstruction before clinical symptoms have manifested.

Advancements of proteomic technologies, such as protein arrays, two-dimensional gels and mass spectrometry (MS) have enabled the identification of biomarker candidates from hundreds to thousands of proteins in a short time using only small sample amounts. We previously developed a targeted multiplex proteomics approach to the relative quantification of >100 proteins based on MS analysis of protein-derived tryptic peptides [12]. Herein, we describe the application of this approach to two independent cohorts: one CAL and one COPD. The aim of this investigation was to find candidate biomarkers for the identification of CAL that could be replicated in a physician-diagnosed COPD cohort.

Materials and methods

Study population

The Burden of Lung Disease (BOLD) study has been described previously [13, 14]. Briefly, the study objective was to estimate COPD prevalence in the general population globally. Participants were randomly selected from people aged ⩾40 years [15]. BOLD subjects were classified as CAL based on airflow limitation, defined using a fixed cut-off of 0.7 for FEV_1/FVC after bronchodilation with 200 µg salbutamol. Slightly more than 15% of BOLD subjects were classified as CAL (figure 1).

The study Tools for Identifying Exacerbations (TIE) in COPD has been described previously [16]. All subjects in TIE have a clinical diagnosis of COPD that was spirometry-verified, by post-bronchodilatory FEV_1/highest of forced or slow vital capacity <0.7 at inclusion in the study. Both studies were approved by the regional ethics review board of Uppsala (Dnr 2006/146 (BOLD) and Dnr 2013/358 (TIE)).

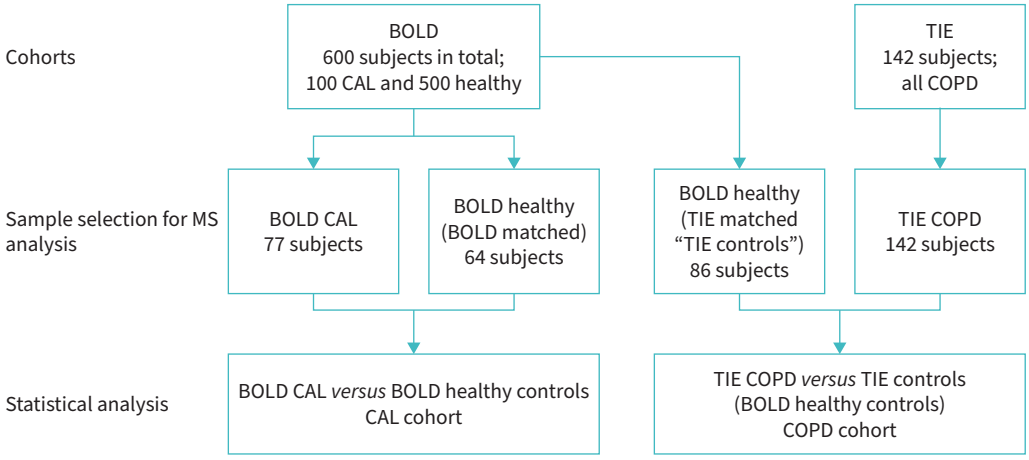

FIGURE 1 A total of 369 samples were selected for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis from a chronic airflow limitation (CAL) (Burden of Obstructive Lung Disease (BOLD)) and a COPD (Tools for Identifying Exacerbations (TIE)) cohort, both from Uppsala, Sweden. Healthy controls were matched to CAL and COPD subjects, with regards to age, sex and smoking status. TIE lacks healthy controls, why a set of healthy controls from BOLD were matched against TIE samples. The TIE study comprises samples collected at an inclusion visit and at follow-up visits. Only inclusion visit samples were included in the LC-MS/MS analysis described herein.

For the study described herein, serum samples from BOLD and TIE participants from Uppsala were used. Only baseline (inclusion) data from the TIE participants was used. All TIE participants were sampled in a stable disease phase, ⩾4 weeks after an exacerbation. We used healthy subjects from BOLD as matched controls for TIE subjects, since the TIE cohort comprises COPD subjects only. Thus, 500 healthy BOLD subjects that did not have CAL were sorted by age, sex and smoking status to identify 150 matched controls for BOLD CAL and TIE COPD subjects. Study participants were characterised as never-, former or current smokers. We categorised healthy subjects and CAL/COPD according to age and smoking status using JMP software (version 15; SAS Institute, Cary, NC, USA) and then selected matched pairs (healthy *versus* CAL/COPD subject). When more than one matched control was available for a CAL/COPD subject, we applied random selection of the available controls. This resulted in 114 perfect matches. To complete the selection of 150 controls perfectly matched for smoking status, the age difference was allowed to increase to ±8 years resulting in 36 additional matches. Altogether, the selection procedure resulted in a group of 150 controls from the BOLD cohort that match with 86 individuals from the TIE and 64 individuals from the BOLD cohort. In total, we selected 77 CAL subjects from BOLD, 142 COPD subjects from TIE baseline visit and 150 healthy controls from BOLD.

Liquid chromatography-tandem MS analysis

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis was done by a selected reaction monitoring (SRM) approach, as described previously [12]. LC-MS/MS methods are described in detail in the supplementary material. Briefly, >100 LC-MS-based biomarker assays were optimised individually to yield a list of 123 peptides representing 100 proteins (the final 123 peptides and the 100 proteins they represent are listed in supplementary table S1). Each single assay consists of one heavy isotope-labelled synthetic peptide and its corresponding endogenous tryptic peptide. Single SRM assays were subsequently combined into a multiple reaction monitoring assay where all heavy peptides were pooled and then added to all clinical samples prior to LC-MS/MS analysis. Peptides were used as surrogates for determining protein marker levels in the clinical serum samples. BOLD and TIE serum samples selected for LC-MS/MS analysis were digested with trypsin, whereafter the specific peptides for each marker were measured by LC-MS/MS. Peptides were identified by co-eluting light- and heavy-labelled transitions in the chromatographic separation. Identified peptide sequences and fragmentation ions, transitions, for each peptide detected, were summed and the ratio between endogenous transitions and internal standard transitions sum was calculated and used as an arbitrary unit for the subsequent statistical analysis.

Statistical analysis

Receiver operating characteristic (ROC) curve, area under the ROC curve (AUC) and random forest analyses were done using R software to evaluate the diagnostic performance of the biomarkers and explore

the best predictors of CAL/COPD. Comparisons between CAL/COPD and controls were performed using Wilcoxon's rank sum test. Further details on statistical methods and analyses are described in the supplementary material.

Results

Study cohort characteristics

In this study, 369 samples were selected for LC-MS/MS analysis from two cohorts: BOLD and TIE. BOLD is a population-based cohort and TIE a physician-diagnosed COPD cohort. The population characteristics for both cohorts are summarised in table 1. In the TIE COPD group, 32% reported ever having received an asthma diagnosis, while the corresponding figure in the BOLD CAL group was 19%. The TIE COPD group had a higher prevalence of current smoking, as well as cumulative smoking history (pack-years) and more often severe airflow obstruction (GOLD 3 and 4) than the BOLD CAL group.

Targeted mass spectrometry analysis of BOLD and TIE cohorts

BOLD and TIE serum samples were prepared as described earlier and analysed using LC-MS/MS. This enabled a relative comparison of peptide signals with CAL and COPD samples *versus* healthy controls and was used as a measure of the level of each peptide, and a surrogate measure of the corresponding protein. Out of a total of 123 peptides included in the MS analysis, 73 peptides representing 61 proteins were successfully detected by LC-MS/MS analysis. ROC curve analysis was done to sort peptides and select the ones showing the greatest AUC. The 10 peptides with the largest AUC are presented in figure 2, where peptides corresponding to histidine-rich glycoprotein (HRG), α_1-acid glycoprotein (AGP1), and α_1-antitrypsin (α1AT) showed the highest performance. We considered an AUC ⩾0.7 as the threshold for being a promising biomarker candidate.

TABLE 1 Characteristics of the population

	BOLD		TIE	
	No CAL	CAL[#]	No COPD[¶]	COPD
Subjects	64	77	86	142
Women	28 (44)	33 (43)	51 (59)	87 (61)
Age years	66±11	65±11	62±7	66±7
BMI kg·m^{-2}	27±4	27±4	27±4	27±5
Current smokers	10 (16)	15 (19)	24 (28)	40 (28)
Former smokers	27 (42)	35 (45)	62 (72)	102 (72)
Never-smokers	27 (42)	27 (35)		
Smoking pack-years	3.0 (0–21)	11 (0–30)	13 (5–26)	34 (23–41)
FEV_1 % predicted	101±14	83±18	97±15	60±18
Educational level				
Elementary	18 (28)	31 (40)	25 (29)	63 (45)
High school	33 (52)	29 (38)	40 (47)	47 (33)
University or college	13 (20)	17 (22)	21 (24)	31 (22)[f]
mMRC ⩾2	1 (1.6)	7 (11)	5 (6.6)	65(46)
GOLD stage[+]				
1	0	48 (62)	0	25 (18)
2	0	24 (31)	0	73 (51)
3	0	5 (6.5)	0	39 (27)
4	0	0	0	5 (3.5)
Ever diagnosed with asthma	7 (11)	15 (19)	10 (12.0)	46 (32)
ICS use	3 (4.7)	14 (18)	5 (5.8)	89 (63)
Heart disease[§]	12 (19)	17 (22)	9 (10)	27 (19)
Hypertension	23 (36)	27 (35)	28 (33)	67 (47)
Diabetes	3 (4.7)	1 (1.3)	4 (4.6)	14 (9.9)[f]

Data are presented as n, n (%), mean±SD or median (interquartile range). BOLD: Burden of Obstructive Lung Disease; TIE: Tools for Identifying Exacerbations; CAL: chronic airflow limitation; BMI: body mass index; FEV_1: forced expiratory volume in 1 s; mMRC: modified Medical Research Council dyspnoea scale; GOLD: Global Initiative for Chronic Obstructive Lung Disease; ICS: inhaled corticosteroid. [#]: defined by presence of FEV_1/forced vital capacity <0.7; [¶]: BOLD controls matched to TIE COPD subjects; [+]: GOLD stages (only in subjects with CAL): 1: FEV_1 ⩾80% predicted, 2: ⩾50% predicted but <80% predicted, 3: ⩾30% predicted but <50% predicted, 4: <30% predicted; [§]: defined as suffering from at least one of angina, myocardial infarction, heart failure or atrial fibrillation; [f]: information missing for one individual.

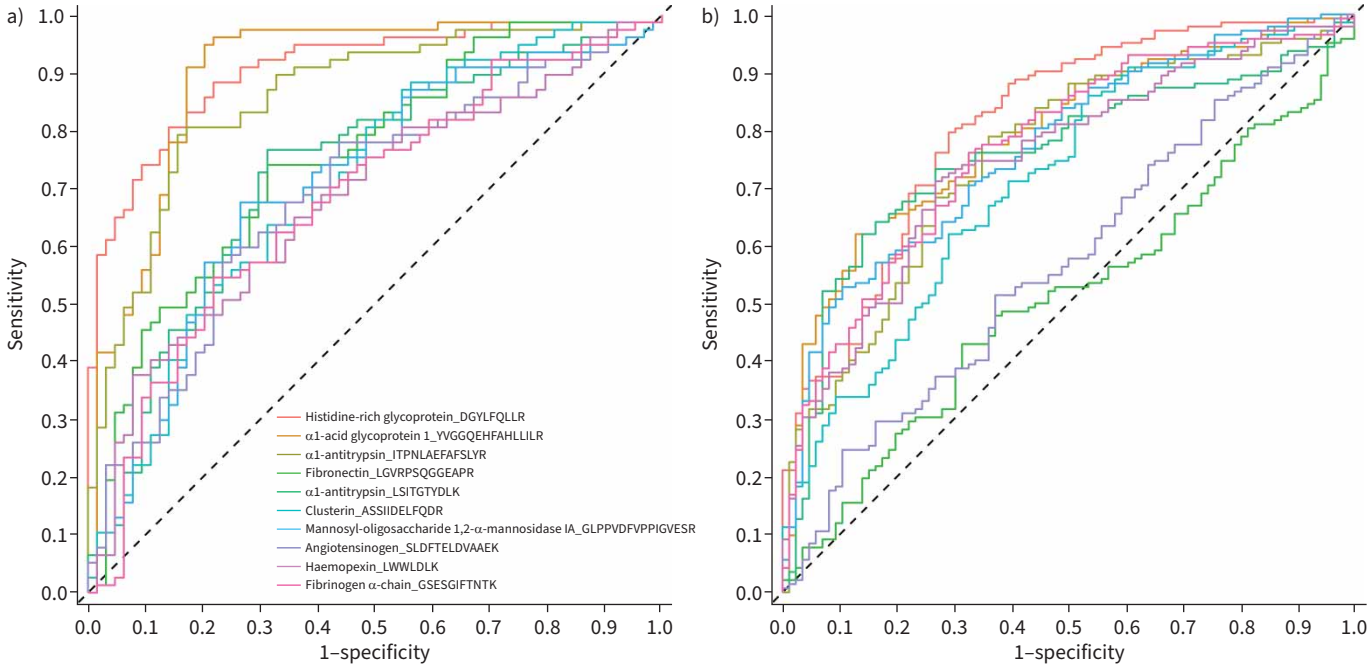

FIGURE 2 Receiver operating characteristic (ROC) curve analysis of selected reaction monitoring assay results. a) Burden of Obstructive Lung Disease (BOLD) chronic airflow limitation serum samples and BOLD matched controls. The 10 peptides with the highest area under the curve from ROC curve analysis of all 73 peptides are presented. b) ROC curve analysis of the same 10 peptides as in (a) from analysis of Tools for Identifying Exacerbations (TIE) COPD serum samples and TIE controls.

As the measurement levels vary across peptides, the values were subsequently transformed to a 0–1 scale to facilitate a boxplot covering the 10 peptide measurements with the greatest AUC (figure 3a), as well as all peptide measurements by group (CAL *versus* controls) for the BOLD cohort (supplementary figure S1). Several peptides showed a higher mean LC-MS/MS signal for CAL compared to controls, where HRG and AGP1 were shown to exhibit the highest ratio: 2.1 and 2.0, respectively (table 2).

LC-MS/MS analysis was also performed with samples from the TIE COPD cohort, using matched BOLD samples as controls (figure 1). Six out of the seven peptides from the BOLD cohort ROC curve analysis with an AUC ⩾0.7 were also shown to have an AUC ⩾0.7 in the TIE cohort ROC curve analysis (table 2). Interestingly, HRG and AGP1 peptides showed the largest AUC for both the BOLD and TIE cohorts (table 2 and supplementary table S2; supplementary table S2 also shows additional peptides identified from the TIE cohort analysis). Overall, several peptides showed an increase in mean LC-MS/MS signal for the COPD group compared to controls (figure 3b and supplementary figure S2), where HRG, AGP1 and α1AT were 1.6-, 1.5- and 1.3-fold higher, respectively, in COPD compared to controls (table 2).

For some proteins, more than one peptide was included in the LC-MS/MS analysis. AGP1 and α1AT were both represented by two peptides. Correlation between peptides representing the most important predictors of CAL (AGP1, HRG and a1AT) are presented in supplementary figure S3.

Smoking and inhaled corticosteroid use

There was no correlation between current smoking and the top three peptides for BOLD CAL or TIE COPD when analysed separately, and the results were consistent for both cohorts also after we adjusted for pack-years (data not shown).

Inhaled corticosteroid (ICS) use varied between groups, ranging from 5% for controls to 18% for BOLD CAL and 63% for TIE COPD. Since ICS use may impact biomarker concentrations, the top 10 biomarkers were also analysed by an adjusted ROC curve to adjust for impact of ICS use. For the BOLD cohort, adjusted (a)AUC for the top three peptides were very similar to the un-adjusted AUC, while aAUC for TIE were slightly higher than AUC (table 2). Thus, the diagnostic performance for CAL and COPD *versus* controls remain after adjustment for ICS.

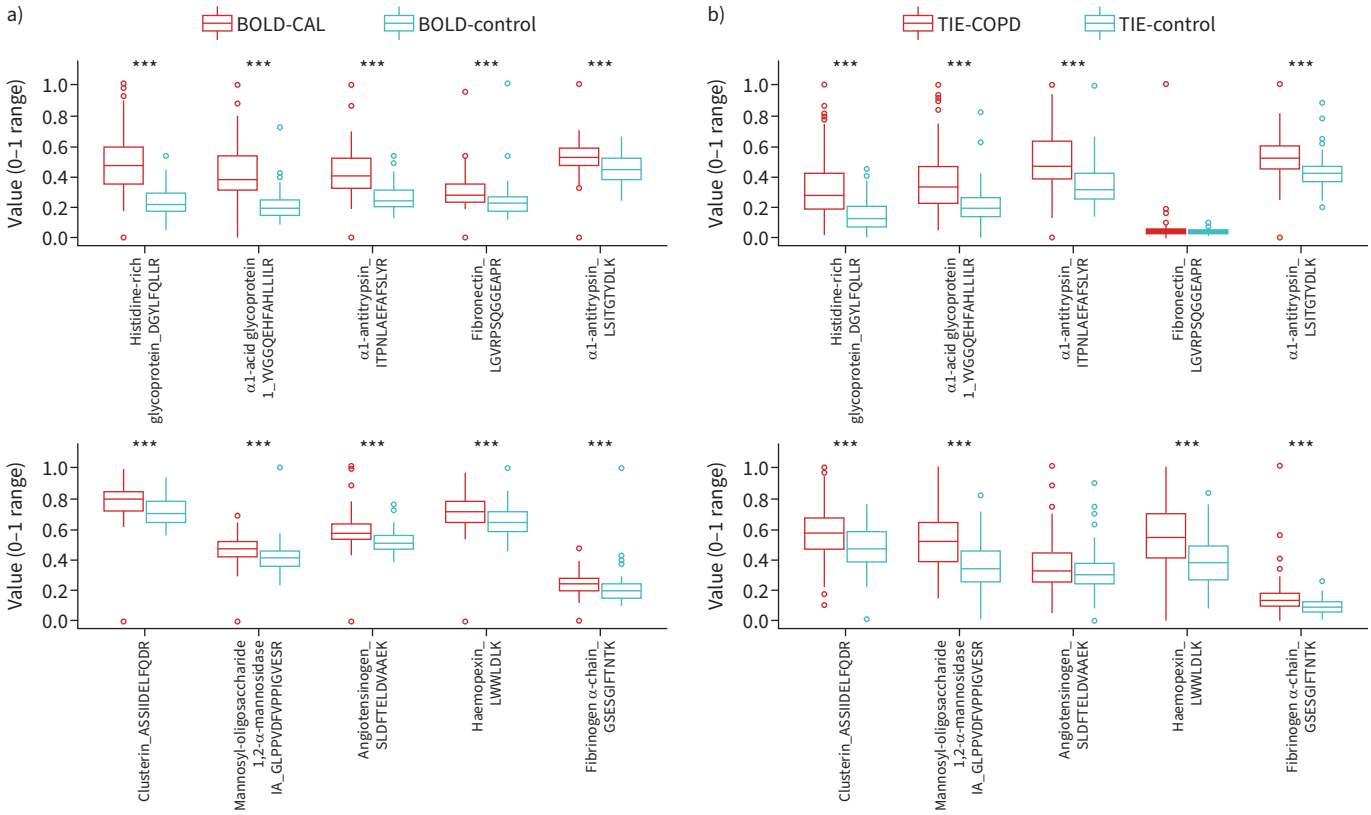

FIGURE 3 Selected reaction monitoring assay results for the 10 peptides with the highest area under the curve (AUC) from analysis of Burden of Obstructive Lung Disease (BOLD) chronic airflow limitation (CAL) *versus* BOLD controls. a) Peptide levels in BOLD CAL and BOLD controls, sorted by AUC; b) peptide levels in Tools for Identifying Exacerbations (TIE) COPD and TIE controls. Data are presented as median (interquartile range (IQR)) and range (no further than 1.5×IQR from the boxes). Outliers are plotted individually. p-values are from Wilcoxon rank sum (asymptotic) tests of the hypothesis that there is no location shift in distribution of peptide levels a) between BOLD-CAL and BOLD-control or b) between TIE-COPD and TIE-control. ***: p<0.001

TABLE 2 Summary of multiple reaction monitoring assay results for Burden of Obstructive Lung Disease (BOLD) chronic airflow limitation (CAL) and Tools for Identifying Exacerbations (TIE) COPD top 10 peptides sorted based on area under the curve (AUC) from BOLD and TIE receiver operating characteristic curve analysis, respectively, along with mean CAL/mean control and mean COPD/mean control ratios, respectively, for the 10 peptides

	Uniprot ID	Short name	Peptide sequence	BOLD CAL			TIE COPD		
				AUC	Ratio	aAUC	AUC	Ratio	aAUC
Histidine-rich glycoprotein[#]	P04196	HRG	DGYLFQLLR	0.91	2.09	0.91	0.81	1.61	0.84
α$_1$-acid glycoprotein 1[#]	P02763	AGP1	YVGGQEHFAHLLILR	0.90	2.02	0.89	0.79	1.50	0.85
α$_1$-antitrypsin[#]	P01009	α1AT	ITPNLAEFAFSLYR	0.86	1.57	0.85	0.75	1.30	0.78
Fibronectin	P02751	FN	LGVRPSQGGEAPR	0.75	1.27	0.60	0.51	1.15	0.59
α$_1$-antitrypsin[#]	P01009	α1AT	LSITGTYDLK	0.73	1.16	0.73	0.75	1.18	0.76
Clusterin[#]	P10909	CLU	ASSIIDELFQDR	0.71	1.08	0.68	0.71	1.10	0.79
Mannosyl-oligosaccharide 1.2-α-mannosidase IA[#]	P33908	Mannosidase 1	GLPPVDFVPPIGVESR	0.71	1.10	0.65	0.77	1.16	0.84
Angiotensinogen	P01019	Serpin A8	SLDFTELDVAAEK	0.69	1.11	0.68	0.57	1.05	0.45
Haemopexin	P02790	HPX	LWWLDLK	0.68	1.08	0.67	0.75	1.13	0.81
Fibrinogen α-chain	P02671	FGA	GSESGIFTNTK	0.68	1.13	0.57	0.77	1.36	0.76

aAUC: AUC adjusted for inhaled corticosteroid use. [#]: AUC ⩾0.7 in both populations.

TABLE 3 Analytical performance based on a multivariate random forest analysis: classification of chronic airflow limitation (CAL) *versus* controls based on random forest

	Reference	
	BOLD CAL	**BOLD control**
Predicted		
BOLD CAL	67	14
BOLD control	9	50

BOLD: Burden of Obstructive Lung Disease.

TABLE 4 Analytical performance based on a multivariate random forest analysis: test performance based on random forest including all 73 peptides detected in the liquid chromatography-tandem mass spectrometry analysis of Burden of Obstructive Lung Disease (BOLD) chronic airflow limitation *versus* BOLD matched controls

	Test performance with 73 peptides
Accuracy	84%
Sensitivity	88%
Specificity	78%
Positive predictive value	83%
Negative predictive value	85%

One sample was removed due to missing data.

Multivariate analysis of MS data

In addition to the univariate AUC method, a multivariate analysis by random forest was done to search for the peptides that were most useful to predict outcome (table 3; CAL *versus* control). The operating characteristics presented in table 4 are from a typical random forest analysis including all 73 peptides that were successfully detected using LC-MS/MS. Correct classification (*i.e.* accuracy) occurred in 84% of cases with a tendency to be better at correctly classifying subjects with CAL (88% sensitivity) than controls (78% specificity). The result of the random forest indicated that three of the peptides clearly stood out as the most important ones for predicting outcome in BOLD (based on mean decrease in Gini; supplementary figure S4A), in order of importance: AGP1 peptide YVGGQEHFAHLLILR, HRG peptide DGYLFQLLR and α1AT peptide ITPNLAEFAFSLYR. For TIE, the three most important peptides were AGP1 peptide YVGGQEHFAHLLILR, HRG peptide DGYLFQLLR, and α1AT peptide LSITGTYDLK (supplementary figure S4B).

To investigate if the set of peptides used for model prediction can be reduced from 73 while maintaining a high accuracy, three separate random forest models were performed on BOLD data, based on the top one, three and 10 peptides as ranked by AUC. Table 5 shows the accuracy of the three random forest models to correctly predict each subject as belonging to BOLD CAL or BOLD controls. The results indicate that maximum accuracy, 85%, is already achieved when only including the top three peptides ranked by AUC: HRG peptide DGYLFQLLR, AGP1 peptide YVGGQEHFAHLLILR and α1AT peptide ITPNLAEFAFSLYR.

TABLE 5 Analytical performance based on a multivariate random forest analysis: accuracy of random forest models based on a combination of different sets of peptides (selection of top one, three and 10 peptides according to Burden of Obstructive Lung Disease area under the curve)

	Top 1 (HRG)	**Top 3 (HRG, AGP1, α1AT)**	**Top 10[#]**
Accuracy	72%	85%	85%

To improve the stability of the accuracy calculation each model was run through a three-fold cross-validation 10 times and the average accuracy was used HRG: histidine-rich glycoprotein; AGP1: α_1-acid glycoprotein; α1AT: α_1-antitrypsin. [#]: table 2.

BOLD versus *TIE*

To explore any differences between BOLD and TIE, a random forest was applied to TIE data (together with their matched controls from BOLD). The results show that the most important peptides were not as clearly separated as for BOLD (supplementary figure S4A and B), although the operating characteristics for a typical random forest analysis of TIE showed 84% accuracy, 90% sensitivity and 74% specificity, which was similar to BOLD.

The outcome for the TIE data was also predicted using the BOLD model. This only achieved an accuracy of 67% (53% sensitivity and 89% specificity). Notably, TIE subjects had lower peptide levels on average than CAL subjects in BOLD.

The 10 peptides with highest AUC were somewhat different based on TIE data compared to BOLD (compare table 2 and supplementary table S2). However, running two separate random forest models on TIE data, one for each set of 10 peptides, showed the accuracy was ∼83% for both models, indicating that the most important peptides are shared between the two datasets.

Discussion

We have used LC-MS/MS to identify potential biomarkers for CAL confirmation and COPD diagnosis, from a panel of 100 biomarker candidates. The biomarkers were identified in a population-based study and subsequently replicated with an independent COPD cohort. The most promising results were seen with AGP1, HRG and α1AT peptides to discriminate both CAL and COPD from control subjects. Within the BOLD cohort, the biomarkers identified herein could discriminate subjects with an FEV_1/FVC <0.7 from subjects with a ratio ⩾0.7, *i.e.* those that did not have an airflow limitation. Since BOLD is a population study where subjects were classified as CAL based on a post-bronchodilatory FEV_1/FVC <0.7 [13], only a few individuals had a physician-diagnosed COPD [17]. Therefore, BOLD subjects with CAL probably represent a group with COPD that did not have manifest clinical symptoms, while all TIE subjects have physician-diagnosed COPD. The findings were stronger in CAL in BOLD subjects than in patients with established COPD from TIE, when applying a cut-off derived from BOLD. The reason is that COPD subjects in TIE had lower peptide levels on average than CAL subjects in BOLD and therefore fewer subjects will be predicted as COPD in TIE, when using a cut-off from BOLD. Altogether, HRG, AGP1 and α1AT may potentially be best used for the assessment of subjects with clinical suspicion of COPD and perhaps also for population screening to look for individuals at risk of developing COPD.

AGP1 is a major acute-phase protein in humans. Its serum concentration is known to increase in response to systemic tissue injury, inflammation or infection [18, 19], but the physiological function of AGP1 in the human body remains unknown. Expression of the AGP1 (or ORM-1) gene is known to be upregulated by glucocorticoids and cytokines interleukin (IL)-1β [20], tumour necrosis factor-α and IL-6 [21]. ORM1 gene expression and plasma levels of AGP1 protein were shown to be increased in acute exacerbations of COPD [22]. Branching of AGP glycans was shown to increase in COPD, which would indicate an inflammatory reaction that differs from a normal acute-phase response, in which there is a decrease in branching of AGP [20].

α1AT is an inhibitor of neutrophil proteases and other pro-inflammatory responses [23] and thus an important regulator of the immune response. α1AT is also a major acute-phase protein and a well-known marker of α1AT deficiency, a genetic condition characterising one distinct COPD endotype [24]. Lower plasma α1AT levels due to genetic deficiency have been shown to be associated with increased risk of COPD exacerbations [25]. However, this genetic condition is rare and unlikely to influence our results. Some studies have shown a higher α1AT plasma protein mean level in COPD patients [26], while other studies failed to see an association between exacerbations or GOLD stage and α1AT serum protein levels [27, 28]. The reason for the discrepancy is possibly due to the small observed actual differences in α1AT level between healthy *versus* COPD groups, where any variation in inflammation at the time of sampling would influence the result. The small actual differences between healthy *versus* COPD groups make it seem logical that α1AT could instead form part of a composite score with inclusion of other relevant biomarkers, rather than a standalone test [27], a concept we believe the study herein has corroborated.

HRG is an abundant blood protein with a multidomain structure. HRG interacts with many ligands and may regulate cell adhesion and migration, fibrinolysis and coagulation, complement activation, immune complex clearance and phagocytosis of apoptotic cells [29]. There are a few reports on HRG in the context of COPD. According to Titz *et al.* [30], HRG levels decrease in early-stage COPD (GOLD stage 1–2 and current smokers) in sputum, in contrast to our findings that show an elevation in HRG levels in CAL.

Additional acute-phase and other proteins identified in this study as upregulated in CAL and COPD were fibronectin, clusterin, mannosyl-oligosaccharide 1,2-α-mannosidase IA, angiotensinogen, haemopexin and fibrinogen. Several acute-phase proteins have previously been proposed as diagnostic biomarkers, such as CRP, α-2-macroglobin, haptoglobin, fibrinogen and IL-6 [31, 32]. However, only a few have made it into clinical diagnostic use where CRP is considered as the most important acute-phase protein [33, 34]. The specificity of CRP for COPD assessment is limited, and additional biomarkers are needed to enable objective identification of COPD and monitoring of disease progression [35, 36]. Assuming inflammation is part of disease onset in COPD, some acute-phase proteins may serve as early indicators of the disease. A potential barrier for using these acute reactant proteins for diagnosing COPD is the disease specificity since inflammation is activated in a plethora of diseases. This could limit the use of acute phase proteins as diagnostic markers of a specific disease. However, the combination of several acute-phase proteins might be a way to deal with this as using three acute-phase proteins increased specificity and sensitivity compared to using one or two markers.

The purpose of this study was to identify biomarkers of CAL and COPD. Such biomarkers would be valuable since there is no curative medication for COPD, early symptoms are subtle, and diagnosis currently only occurs after lung impairment has manifested. This may enable diagnosis at a point when nonpharmacological interventions like smoking cessation and avoiding hazardous environments could prevent disease progression.

A strength of the study is that we have tested our entire biomarker panel on a population cohort, as well a physician-diagnosed COPD cohort. This enabled us show that the most useful biomarkers for CAL were also the best predictors of COPD. The fact that we could demonstrate that the biomarkers worked slightly better in the population-based cohort suggests that these biomarkers should be useful if they are used in a screening setting, compared to if our findings would have been based only on a COPD cohort. A limitation of the SRM approach described was that, for some proteins in our initial panel of biomarkers, it was not possible to identify unique, non-homologous peptides that would give a satisfactory signal in the LC-MS/MS analysis. Thus, some proteins could not be included in the study. Another limitation is that some of the subjects in BOLD may have CAL because of other causes than COPD, such as asthma [17], and that we used a lower bronchodilator dose than generally recommended by the American Thoracic Society (ATS)/European Respiratory Society (ERS) 2005 guidelines [37]. Other potential limitations are that the assessments were done at only one time point and that we only assessed biomarkers in the blood and did not include biomarkers from sputum. Repeated measurements would have given us more information on the association between biomarkers and disease progression and/or lung function trajectories, and it is possible that airway biomarkers would be better related to CAL and/or COPD than circulating molecules.

In summary, we have identified AGP1, HRG and α1AT as three promising biomarker candidates capable of distinguishing CAL, identified from a general population, and replicated our findings with a physician-diagnosed COPD patient cohort. All three biomarkers by themselves exhibit very good clinical performance with high sensitivity and specificity, but interestingly seem to complement each other since an even better test performance was observed when combining the three biomarkers. The combined performance suggests these biomarkers may be useful for identifying individuals with CAL and COPD.

Provenance: Submitted article, peer reviewed.

Author contributions: M. Molin, A. Incamps, M. Lemasson, M. Andersson, E. Pertsinidou, M. Högman, K. Lisspers, B. Ställberg, A. Sjölander, A. Malinovschi and C. Janson participated in the study design, interpretation of the data, drafting the manuscript and approved the final version of the manuscript. M. Molin, A. Incamps, M. Andersson, A. Sjölander, A. Malinovschi and C. Janson prepared the first draft of the manuscript. A. Incamps and M. Lemasson performed the mass spectrometry analysis. M. Molin, A. Sjölander, M. Andersson, A. Malinovschi and C. Janson performed the statistical analysis. C. Janson and A. Malinovschi take responsibility for the integrity of the entire work, from inception to published article.

Conflict of interest: M. Molin, A. Incamps, M. Lemasson, E. Pertsinidou, M. Andersson and A. Sjölander are or were employed by Thermo Fisher Scientific. M. Högman, K. Lisspers, B. Ställberg, A. Malinovschi and C. Janson declare no competing interests in relation to this study.

Support statement: This study was supported by the Uppsala-Örebro Regional Research Council, The Swedish Heart–Lung Foundation, The Swedish Heart and Lung Association, and the Swedish Research Council. GlaxoSmithKline supported the BOLD study with an unrestricted grant. Thermo Fisher Scientific performed the

mass spectrometry analyses in a research collaboration agreement with Uppsala University. Funding information for this article has been deposited with the Crossref Funder Registry.

Ethics statement: Study participants gave their written and informed consent. Both the BOLD and TIE studies were approved by the regional ethics review board of Uppsala, Sweden: Dnr 2006/146 (BOLD) and Dnr 2013/358 (TIE).

References

1 World Health Organization (WHO). WHO Global Health Estimates. 2019. 2020. Available from: www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
2 Halpin DMG, Criner GJ, Papi A, *et al.* Global initiative for the diagnosis, management, and prevention of chronic obstructive lung disease. The 2020 GOLD science committee report on COVID-19 and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2021; 203: 24–36.
3 Childs BG, Durik M, Baker DJ, *et al.* Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nat Med* 2015; 21: 1424–1435.
4 Chilosi M, Carloni A, Rossi A, *et al.* Premature lung aging and cellular senescence in the pathogenesis of idiopathic pulmonary fibrosis and COPD/emphysema. *Transl Res* 2013; 162: 156–173.
5 Anthonisen NR, Skeans MA, Wise RA, *et al.* The effects of a smoking cessation intervention on 14.5-year mortality: a randomized clinical trial. *Ann Intern Med* 2005; 142: 233–239.
6 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. 2022 GOLD report. 2022. Available from: www.goldcopd.org.
7 Serban KA, Pratte KA, Bowler RP. Protein biomarkers for COPD outcomes. *Chest* 2021; 159: 2244–2253.
8 Agusti A, Sin DD. Biomarkers in COPD. *Clin Chest Med* 2014; 35: 131–141.
9 Papi A, Bellettato CM, Braccioni F, *et al.* Infections and airway inflammation in chronic obstructive pulmonary disease severe exacerbations. *Am J Respir Crit Care Med* 2006; 173: 1114–1121.
10 Gabay C, Kushner I. Acute-phase proteins and other systemic responses to inflammation. *N Engl J Med* 1999; 340: 448–454.
11 Eapen MS, Myers S, Walters EH, *et al.* Airway inflammation in chronic obstructive pulmonary disease (COPD): a true paradox. *Expert Rev Respir Med* 2017; 11: 827–839.
12 Incamps A, Saez-Boiteau C, Tiede SL, *et al.* Impact of biological matrix on inflammatory protein biomarker quantification based on targeted mass spectrometry. *Bioanalysis* 2018; 10: 1383–1399.
13 Buist AS, Vollmer WM, Sullivan SD, *et al.* The burden of obstructive lung disease initiative (BOLD): rationale and design. *COPD* 2005; 2: 277–283.
14 Buist AS, McBurnie MA, Vollmer WM, *et al.* International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet* 2007; 370: 741–750.
15 Danielsson P, Ólafsdóttir IS, Benediktsdóttir B, *et al.* The prevalence of chronic obstructive pulmonary disease in Uppsala, Sweden – the Burden of Obstructive Lung Disease (BOLD) study: cross-sectional population-based study. *Clin Respir J* 2012; 6: 120–127.
16 Högman M, Sulku J, Ställberg B, *et al.* Global initiative for chronic obstructive lung disease reclassifies half of COPD subjects to lower risk group. *Int J Chron Obstruct Pulmon Dis* 2017; 2018: 165–173.
17 Bröström E, Jõgi R, Gislason T, *et al.* The prevalence of chronic airflow obstruction in three cities in the Nordic-Baltic region. *Respir Med* 2018; 143: 8–13.
18 Fournier T, Medjoubi NN, Porquet D. Alpha-1-acid glycoprotein. *Biochim Biophys Acta* 2000; 1482: 157–171.
19 Ceciliani F, Lecchi C. The immune functions of α1 acid glycoprotein. *Curr Protein Pept Sci* 2019; 20: 505–524.
20 Van Den Heuvel MM, Poland DC, De Graaff CS, *et al.* The degree of branching of the glycans of alpha$_1$-acid glycoprotein in asthma. A correlation with lung function and inflammatory parameters. *Am J Respir Crit Care Med* 2000; 161: 1972–1978.
21 Baumann H. Hepatic acute phase reaction *in vivo* and *in vitro*. *In Vitro Cell Dev Biol* 1989; 25: 115–126.
22 Shi L, Zhu B, Xu M, *et al.* Selection of AECOPD-specific immunomodulatory biomarkers by integrating genomics and proteomics with clinical informatics. *Cell Biol Toxicol* 2018; 34: 109–123.
23 Gadek JE, Fells GA, Zimmerman RL, *et al.* Antielastases of the human alveolar structures. Implications for the protease-antiprotease theory of emphysema. *J Clin Invest* 1981; 68: 889–898.
24 Gooptu B, Ekeowa UI, Lomas DA. Mechanisms of emphysema in α_1-antitrypsin deficiency: molecular and cellular insights. *Eur Respir J* 2009; 34: 475–488.
25 Ingebrigtsen TS, Marott JL, Rode L, *et al.* Fibrinogen and α1-antitrypsin in COPD exacerbations. *Thorax* 2015; 70: 1014–1021.
26 Janciauskiene S, DeLuca DS, Barrecheguren M, *et al.* Serum levels of alpha1-antitrypsin and their relationship with COPD in the general Spanish population. *Arch Bronconeumol* 2020; 56: 76–83.
27 Ellis P, Turner A. What do alpha-1 antitrypsin levels tell us about chronic inflammation in COPD? *Arch Bronconeumol* 2020; 56: 72–73.

28 Nuñez A, Marras V, Harlander M, *et al.* Association between routine blood biomarkers and clinical phenotypes and exacerbations in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2020; 15: 681–690.

29 Poon IK, Patel KK, Davis DS, *et al.* Histidine-rich glycoprotein: the Swiss Army knife of mammalian plasma. *Blood* 2011; 117: 2093–2101.

30 Titz B, Sewer A, Schneider T, *et al.* Alterations in the sputum proteome and transcriptome in smokers and early-stage COPD subjects. *J Proteomics* 2015; 128: 306–320.

31 Jain S, Gautam V, Naseem S. Acute-phase proteins: as diagnostic tool. *J Pharm Bioallied Sci* 2011; 3: 118–127.

32 Schrödl W, Büchler R, Wendler S, *et al.* Acute phase proteins as promising biomarkers: perspectives and limitations for human and veterinary medicine. *Proteomics Clin Appl* 2016; 10: 1077–1092.

33 de Torres JP, Cordoba-Lanus E, López-Aguilar C, *et al.* C-reactive protein levels and clinically important predictive outcomes in stable COPD patients. *Eur Respir J* 2006; 27: 902–907.

34 Dahl M, Vestbo J, Lange P, *et al.* C-reactive protein as a predictor of prognosis in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2007; 175: 250–255.

35 Leung JM, Sin DD. Biomarkers in airway diseases. *Can Respir J* 2013; 20: 180–182.

36 Paone G, Leone V, Conti V, *et al.* Blood and sputum biomarkers in COPD and asthma: a review. *Eur Rev Med Pharmacol Sci* 2016; 20: 698–708.

37 Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.