Accepted Manuscript

Sputum proteomic signature of gastro-oesophageal reflux in patients with severe asthma

Tariq K, MRCP, Schofield JPR, PhD, Nicholas BL, PhD, Burg D, PhD, Brandsma J, PhD, Bansal AT, PhD, Wilson SJ, PhD, Lutter R, PhD, Fowler SJ, MD, Bakke PS, MD, Caruso M, PhD, Dahlen B, PhD, Horváth I, MD, Krug N, MD, Montuschi P, MD, Sanak M, PhD, Sandström T, PhD, Geiser T, MD, Pandis I, PhD, Sousa AR, PhD, Adcock IM, PhD, Shaw DE, MD, Auffray C, PhD, Howarth PH, DM, Sterk PJ, PhD, Chung KF, MD, Skipp PJ, PhD, Dimitrov B, PhD, Djukanović R, MD, the U-BIOPRED Study Group

PII: S0954-6111(19)30044-7

DOI: https://doi.org/10.1016/j.rmed.2019.02.008

Reference: YRMED 5632

To appear in: *Respiratory Medicine*

Received Date: 29 November 2018

Revised Date: 30 January 2019

Accepted Date: 4 February 2019

Please cite this article as: K T, JPR S, BL N, D B, J B, AT B, SJ W, R L, SJ F, PS B, M C, B D, I H, N K, P M, M S, T S, T G, I P, AR S, IM A, DE S, C A, PH H, PJ S, KF C, PJ S, B D, R D, the U-BIOPRED Study Group, Sputum proteomic signature of gastro-oesophageal reflux in patients with severe asthma, *Respiratory Medicine*, https://doi.org/10.1016/j.rmed.2019.02.008.

Sputum proteomic signature of gastro-oesophageal reflux in patients with severe asthma

Tariq K, MRCP [1,3#], Schofield JPR, PhD [1,2#], Nicholas BL, PhD [1,3], Burg D, PhD [1, 2], Brandsma J, PhD [1], Bansal AT, PhD [4], Wilson SJ, PhD [1], Lutter R, PhD [5,6], Fowler SJ, MD [8], Bakke PS, MD [10], Caruso M, PhD [11], Dahlen B, PhD [12], , Horváth I, MD [13], Krug N, MD [14], Montuschi P, MD [15], Sanak M, PhD [16], Sandström T, PhD [17], Geiser T, MD [18], Pandis I, PhD [19], Sousa AR, PhD [20], Adcock IM, PhD [21], Shaw DE, MD [9], Auffray C, PhD [22], Howarth PH, DM [1,3], Sterk PJ, PhD [6], Chung KF, MD [7], Skipp PJ, PhD [2], Dimitrov B, PhD [1], [§]Djukanović R, MD [1,3], and the U-BIOPRED Study Group

[#]: contributed equally as first authors.

[1]NIHR Southampton Respiratory Biomedical Research Centre, University Hospital Southampton, Southampton, UK

[2]Centre for Proteomic Research, University of Southampton, Highfield, Southampton, UK

[3] Clinical Experimental Sciences unit, Faculty of Medicine, University of Southampton, University Hospital Southampton, South Academic Block, Southampton, UK

[4]Acclarogen Ltd, Cambridge, UK

[5]AMC, Department of Experimental Immunology, University of Amsterdam, Amsterdam, The Netherlands

[6]AMC, Department of Respiratory Medicine, University of Amsterdam, Amsterdam, The Netherlands

[7]Airways Disease, National Heart and Lung Institute, Imperial College, London & Royal Brompton NIHR Biomedical Research Unit, London, United Kingdom

[8]Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester and University Hospital of South Manchester, NHS Foundation Trust, Manchester UK

[9]Respiratory Research Unit, University of Nottingham, Nottingham, UK

[10] Department of Clinical Science, University of Bergen, Bergen, Norway

[11]Dept. of Clinical and Experimental Medicine Hospital University, Policlinico-Vittorio Emanuele, University of Catania, Catania, Italy.

[12]Division of Respiratory Medicine and Allergy, Department of Medicine, Karolinska University Hospital Huddinge, Stockholm, Sweden.

[13]Dept. of Pulmonology, Semmelweis University, Budapest, Hungary

[14]Fraunhofer Institute for Toxicology and Experimental Medicine Hannover, Hannover, Germany.

[15]Dept. of Pharmacology, Faculty of Medicine, Catholic University of the Sacred Heart, Rome, Italy

[16]Division of Molecular Biology and Clinical Genetics, Medical College, Jagiellonian University Medical College, Krakow, Poland

[17]Dept. of Medicine, Dept of Public Health and Clinical Medicine Respiratory Medicine Unit, Umeå University, Umeå, Sweden

[18] University Hospital Bern, Bern, Switzerland

[19]Data Science Institute, Imperial College, London, UK

[20]Respiratory Therapeutic Unit, GSK, Stockley Park, UK

[21]Cell and Molecular Biology Group, Airways Disease Section, National Heart and Lung Institute, Imperial College London, Dovehouse Street, London, UK

[22] European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France

§ Corresponding author

Address for correspondence:

Prof Ratko Djukanovic, mailpoint 810, Level F, Sir Henry Wellcome Laboratories, South Block, Southampton University Hospital, Southampton SO16 6YD, UK. **tel**: + 44 (0) 238 120 4195.

Email: rd1@soton.ac.uk

Abstract

Gastro-oesophageal reflux disease (GORD) has long been associated with poor asthma control without an established cause-effect relationship.

610 asthmatics (421 severe/88 mild-moderate) and 101 healthy controls were assessed clinically and a subset of 154 severe asthmatics underwent proteomic analysis of induced sputum using untargeted mass spectrometry, LC-IMS-MS^{E}. Univariate and multiple logistic regression analyses (MLR) were conducted to identify proteins associated with GORD in this cohort.

When compared to mild/moderate asthmatics and healthy individuals, respectively, GORD was three- and ten-fold more prevalent in severe asthmatics and was associated with increased asthma symptoms and oral corticosteroid use, poorer quality of life, depression/anxiety, obesity and symptoms of sino-nasal disease. Comparison of sputum proteomes in severe asthmatics with and without active GORD showed five differentially abundant proteins with described roles in anti-microbial defences, systemic inflammation and epithelial integrity. Three of these were associated with active GORD by multiple linear regression analysis: Ig lambda variable 1-47 (p=0·017) and plasma protease C1 inhibitor (p=0·043), both in lower concentrations, and lipocalin-1 (p=0·034) in higher concentrations in active GORD.

This study provides evidence which suggests that reflux can cause subtle perturbation of proteins detectable in the airways lining fluid and that severe asthmatics with GORD may represent a distinct phenotype of asthma.

Introduction

Asthma is a disease of varying severity with complex underlying mechanisms. Its pathological features have been studied extensively, including in patients with severe disease, [1] but the roles of known and suspected triggers of asthma remain poorly understood. Amongst these is gastro-oesophageal reflux disease (GORD), a co-morbidity widely associated with asthma. Based on history, its prevalence is estimated to be as high as 80%,[2] which is significantly higher than in the general population, while between 32 and 84% of asthmatics have abnormal acid reflux demonstrated by pH studies,[3-5] although a substantial proportion do not have typical symptoms.[3] The higher prevalence of GORD in asthmatics has long been viewed as a risk factor, with evidence of a two-fold increase new diagnosis of asthma and respiratory symptoms in patients with persistent nocturnal reflux[6] and a five-fold increased risk of exacerbations.[7] Our recent assessment of participants in the U-BIOPRED (**U**nbiased **BIO**markers for the **P**rediction or **RE**spiratory **D**isease Outcomes) project confirmed GORD as a significant co-morbidity in severe asthma,[8] in keeping with other studies.[9,10] However, such associations do not necessarily imply a causal relationship between GORD and severe asthma; alternatively, GORD could be the result of severe asthma due to altered lung mechanics, such as hyperinflation, complicated by increased weight, obesity and asthma drugs, which are all associated with GORD.[11]

Proton pump inhibitors (PPI) are effective at controlling GORD symptoms like heartburn[12] but are variably effective at improving asthma control; the same is true for fundoplication which physically blocks reflux.[13] This is the case even when acid reflux is confirmed by pH monitoring.[14] Such variability in response could be due to sub-optimal patient selection. Effectiveness could be improved if biomarkers were available to demonstrate that some of the gastric refluxate is inhaled and that this impacts on the underlying asthma pathobiology. Using this argument as the rationale for the current study, we hypothesised that the airways of severe asthmatics with active GORD are exposed to oropharyngeal refluxate by inhalation into the lower airways where it causes measurable biological effects. To test this hypothesis, we studied more than 240 participants in the U-BIOPRED project with good quality induced sputum samples and applied to their sputum supernatant a state of the art, quantitative liquid chromatography and untargeted mass spectrometry analysis, LC/MS^{E}. Using multiple logistic regression, we then identified the proteins associated with GORD in patients with severe asthma.

METHODS

Study design

U-BIOPRED is a prospective, multicentre cohort study involving sixteen clinical centres in eleven European Union countries, recruiting healthy and asthmatic participants according to pre-set criteria for clinical stratification as previously published.[8] For the purpose of this study, the following clinical data were evaluated: diagnosis of GORD as recommended by standard guidelines[15a] and its activity at the time of assessment, smoking history, atopy, oral corticosteroid (OCS) treatment, exacerbation frequency in the past year, standardised disease activity questionnaires (short and full version of the Asthma Control Questionnaire [ACQ5 and ACQ7), Asthma Quality of Life Questionnaire [AQLQ], Hospital Anxiety and Depression Scores [HADS], Sino-Nasal Outcome Test [SNOT-20], Epworth sleep score [ESS]), spirometry, and exhaled nitric oxide. All participants were asked to provide sputum induced by nebulised saline, usually on the day of clinical assessment; if samples failed the quality criteria or induction was unsuccessful, sampling was repeated within one week.

The study received ethics approval in all the countries involved and all participants provided written informed consent.

Cohort description

The recruitment criteria have been reported previously.[8] A total of 610 adult participants were stratified into four groups: Group A (n=311) - severe asthmatics on high-dose inhaled corticosteroids (ICS) ≥1000 µg fluticasone propionate (or equivalent), with no smoking in the past year and <5 pack-year smoking history, Group B (n=110) - severe asthmatics defined as for Group A but with either current or past (at least 5 pack-year) smoking history, Group C (n=88) – mild-moderate asthmatics with controlled or partially controlled asthma (defined by GINA) using <500 µg fluticasone propionate ICS (or equivalent) with no smoking in the past year and <5 pack-year smoking history, and Group D (n=101) - healthy individuals with no chronic respiratory disease, pre-bronchodilator FEV_1≥80% of predicted and non-smoking for ≥1 year or ex-smokers with a smoking history of <5 pack years.

The groups were further stratified into subgroups by previous physician-made diagnosis of GORD (ALL-GORD subgroup), i.e. all participants with a diagnosis of GORD and/or on treatment with anti-reflux medication) and participants who, at the time of assessment, had symptoms of GORD (ACTIVE-GORD), and those without a history of GORD and not on medication for GORD (NO-GORD).

Sample collection and analysis

Induced sputum was acquired and processed using U-BIOPRED standard operating procedures[16], using dithioerythritol (DTE) as a mucolytic to obtain supernatant for mass spectrometric analysis and cytospins for inflammatory cell counts[17].

Mass spectrometry

For full details of the mass spectrometric analysis, data curation, protein identity searching, data filtering and normalisation see the online supplement and our previous publication[18]. Samples were analysed in duplicate by LC-IMS-MS^E on a Waters G2S high definition mass spectrometer coupled to a nanoAcquity UPLC system. Database searches were performed using a custom package (Regression tester) based upon executable files from ProteinLynx Global Server 3·0 (Waters) and searched against the Uniprot human reference database (20/11/2014) with added sequence information for internal standards. Quantity was estimated in absolute amounts using the Hi 3 method.

Statistical analysis

Clinical and demographic data were analysed by parametric and non-parametric tests following assessment of distribution using the Shapiro-Wilk normality test by GraphPad Prism (version 6·0 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com) and SPSS (IBM SPSS Statistics for Windows, Version 22·0. Armonk, NY, USA). Comparisons of protein profiles defined by MS^E analysis were restricted to the severe asthma groups A and B in order to avoid confounding effects of disease severity. Comparison between protein profiles in ACTIVE-GORD and NO-GORD was the primary endpoint while that between ALL-GORD and NO-GORD participants was secondary. Feature selection for univariate logistic regression (ULR) of proteins predictive of ACTIVE-GORD and ALL-GORD in severe asthma (groups A and B) was performed by selecting proteins with differential concentrations in ACTIVE-GORD and ALL-GORD compared with NO-GORD subgroups (Mann–Whitney U test, $p \leq 0{\cdot}1$, without adjustment for multiple testing and with p-values raw and unadjusted). Analysis was limited to proteins present in ≥60% of participants to counter factors such as missingness which can adversely affect mass spectrometry analysis (see supplement and previous publication[18] for rationale). The proteins shown by ULR to be associated with GORD (with $p \leq 0{\cdot}05$) were selected for multiple logistic regression (MLR) conducted with stepwise backward selection (adjusted for smoking and oral corticosteroid use) to rule out weak associations and select an efficient model of GORD. All regression analyses were conducted in SPSS.

Role of the funding source

The U-BIOPRED study was funded by the European Union Innovative Medicines Initiative (IMI) and Dr Kamran Tariq's salary was funded by a fellowship with the NIHR Southampton Biomedical Research Centre. Neither funders had any role in the study design, collection analysis and interpretation of data, writing of this manuscript, or its submission for publication.

Table 1.Demographic and clinical features of U-BIOPRED cohort groups A, B, C and D.

	Severe asthma (Group A, non-smokers)				Severe asthma (Group B, smokers)				Mild/Moderate asthma (Group C)				Healthy (Group D)			
	All in group	**NO-GORD**	**ALL-GORD**	**ACTIVE-GORD**	**All in group**	**NO-GORD**	**ALL-GORD**	**ACTIVE-GORD**	**All in group**	**NO-GORD**	**ALL-GORD**	**ACTIVE-GORD**	**All in group**	**NO-GORD**	**ALL-GORD**	**ACTIVE-GORD**
N (%)	311 (50·98%)	142 (45·7%)	169 (54·3%)	103 (33·1%)	110 (18·03%)	37 (33·6%)	73 (66·4%)	51 (46·4%)	88 (14·42%)	72 (81·8%)	16 (18·2%)	10 (11·4%)	101 (16·55%)	96 (95%)	5 (5%)	3 (2·97%)
Age (Yrs)	53 (43-62)	51 (41-60)	55 (43·5-62)	53 (42-62)	55 (48-61·3)	53 (47-63)	55 (49-61)	54 (48-63)	42·5 (28-52·8)	38·5 (26·3-52)	46 (35·5-61)	46·5 (34·5-52)	34 (27-49)	34 (26·3-48·8)	54 (45-61·5)	51 (41-54)**
SEX -M/F (%)	106/205 (34/66)	58/84	48/121*	32/ 71	54/56 (49/51)	23/14	31/42	22/29	44/44 (50/50)	38/34	6/10	4/6	62/39 (61/39)	58/38	4/1	3/0
Atopy POS/NEG (%POS)	213 /59 (78·3%)	104/24 (38·2%)	109/35 (40·1%)	68/21 (25%)	62/25 (71·3%)	23/8 (26·4%)	39/17 (44·8%)	25/12 (28·7%)	72 /6 (92·3%)	59/4 (75·6%)	13/2 (16·7%)	8/1 (10·3%)	36/42 (46·2%)	34/41 (43·6%)	2/1 (2·6%)	1/0 (1·3%)
BMI	27·7 (24·5-33·6)	26 (23·8-31·3)	29·7 (25·2-34·5)**	28·93 (24·6-34·4)*	28·88 (25·2-32·6)	27·3 (24·3-31·3)	29·56 (25·9-33·4)*	29·24 (25-33·7)	24·85 (23-28·9)	24·55 (21·9-27·8)	28·52 (24·6-32·8)**	28·52 (25·9-34·2)**	24·69 (22·7-27·5)	25·12 (22·8-27·5)	23·9 (20-26·6)	23·9 (20·1-28·9)
Smoking (Pack-years)	0	0	0	0	17·38 (10-26)	20 (10-23·6)	16·5 (10·5-30)	16·4 (10-25)	0	0	0	0	0	0	0	0
On oral steroids (%)	122 (n=311) 39·2%	43 (35·2%)	79 (64·8%) **	44 (36·1%)*	36 (n=110) 32·7%	11 (30·5%)	25 (69·4%)	16 (44·4%)	0	0	0	0	0	0	0	0
Exacerbations in last 12 months	2 (2-4)	2 (2-3)	3 (2-4)	3 (1-4)	2 (1-4)	2 (2-4.3)	2 (1-4)	N/A	0	0	0	0	0	0	0	0
FEV_1 (% pred)	67·49 (50·7-84·9)	66·4 (47·7-85·7)	67·7 (52·5-84·1)	67·36 (54·9-81·5)	65·86 (51·8-77·8)	65·68 (52·8-82·1)	67·68 (51·7-76·3)	71·56 (51·3-78·8)	91·74 (75·9-101·7)	92·38 (74·5-99·7)	89·35 (83·9-105·3)	102·8 (82·3-106·1)	102·14 (93·6-110·7)	102·4 (94-111)	99 (86·8-107·5)	105·3 (99-109·8)
FVC (% pred)	87·22 (19·6)	86·1 (19·4)	88·2 (19·8)	87·45 (19·4)	89·72 (18·2)	91·04 (19)	89·05 (17·8)	88·72 (16·8)	104·4 (18·9)	103·4 (19·8)	109·7 (13·2)	108·7 (15·7)	107·8 (13·4)	107·9 (13·7)	104·8 (100·9-112·6)	104·8 (100·7-112·4)
Exhaled NO	26·5 (15·5-47·6)	26 (16-46)	27 (15-48·5)	31 (18·3-48·5)	23·5 (12-43·5)	29·5 (14·1-41·3)	21·5 (11·9-51)	17·75 (10·3-51)	25 (18-55)	27·5 (18-60)	20·25 (13·5-37)	22·5 (9·5-35)	19·25 (13·1-29)	19 (13·5-29)	24 (11·8-41·5)	24 (12·5-56)

Sputum-Macrophage %	26·24 (10·4-48·7)	24·6 (9·8-38·6)	29·7 (12-52·1)	39·86 (11·5-52·6)	33·73 (15·1-48·8)	30·25 (15·3-44·4)	34·76 (13·6-50·9)	34·32 (17·5-49·7)	44·01 (35·8-68·1)	44·58 (36-68·8)	36·16 (18·2-40·4)	29·28 (18·2-40·4)	58·92 (37·5-76·7)	58·1 (37·4-77·1)	68·79 (62·2-75·4)	75·42 (75·4-75·4)
Sputum-Eosinophil %	2·75 (0·4-20·5)	2·6 (0·4-25·5)	2·9 (0·37-17·5)	3·44 (0·4-14·7)	3·81 (0·7-13·7)	4·89 (1·6-19·5)	2 (0·4-12·3)	1.9 (0·4-6·2)	0·78 (0·2-3·4)	0·79 (0·2-3·4)	0·21 (0-3·4)	1·68 (0-3·4)	0 (0-0·3)	0 (0-0·4)	0	0
Sputum-Neutrophil %	53·69 (32·5-76·4)	56·6 (37·2-78·9)	52·3 (30·6-73·5)	50·65 (34·2-74·4)	55·1 (35-65·9)	55·52 (34·8-66·8)	54 (35-65·3)	57·3 (37·4-74·6)	41·71 (23·7-63·3)	38·78 (22·9-59·3)	76·4 (63·6-76·7)*	76·56 (76·4-76·7)*	39·56 (20·9-61·5)	39·81 (21·2-62·2)	28·6 (20·8-36·5)	20·76 (20·7-20·7)

Continued table 1.

Note: Data from each group are shown for the whole group (All in group), for the subgroup of participants with no history of GORD (NO-GORD), for the subgroup with a history/diagnosis/current treatment of GORD (ALL-GORD) and for those with typical symptoms of GORD at the time of assessment (ACTIVE-GORD).

*= p<0·05 and **= p<0·01 for the comparison between participants in the NO-GORD category as compared with ACTIVE-GORD or ALL-GORD, respectively. All values are shown as median (range) or mean (SD) depending on type of data distribution.

BMI = Body mass index (Kg/m^2). FEV_1 = Forced expiratory volume in 1 second, FVC = Forced vital capacity. Atopy results were available for 515 individuals based on skin prick tests and/or RAST. Group A - 272, Group B - 87, Group C – 78 and Group D - 78. Pos = positive for atopy, Neg = negative for atopy. Exhaled NO = exhaled nitric oxide levels in parts per million. % = percentage.

Table 2. Patient reported outcomes in the U-BIOPRED cohort groups A, B, C and D**.**

	Severe asthma (Group A, non-smokers)				Severe asthma (Group B, Smokers)				Mild/Moderate asthma (Group C)				Healthy (Group D)			
	All in group	NO-GORD	ALL-GORD	ACTIVE-GORD	All in group	NO-GORD	ALL-GORD	ACTIVE-GORD	All in group	NO-GORD	ALL-GORD	ACTIVE-GORD	All in group	NO-GORD	ALL-GORD	ACTIVE-GORD
ACQ5	2·2 (1·4-3·2)	2·2 (1·2-3)	2·4 (1·5-3·3)	2·4 (1·8-3·2)	2·2 (1·4-3)	1·9 (0·6-2·8)	2·5 (1·8-3·2)**	2·8 (1·8-3·4)**	0·8 (0·3-1·4)	0·8 (0·3-1·4)	1 (0·3-1·6)	0·8 (0-1·7)	N/A	N/A	N/A	N/A
ACQ7	2·71 (1·7-3·6)	2·57 (1·6-3·4)	2·8 (1·7-3·7)	2·86 (2-3·7)	2·57 (1·7-3·4)	2·29 (1.2-3)	2·79 (1·8-3·9)*	2·93 (1·9-3·9)*	1·0 (0·4-1·6)	1·0 (0·4-1·5)	1·14 (0·4-1·7)	0·93 (0-1·7)	N/A	N/A	N/A	N/A
AQLQ	4·51 (3·6-5·4)	4·59 (3·8-5·5)	4·31 (3·2-5·4)	4.31 (3.5-5.3)	4·44 (3·5-5·3)	5·03 (4·3-6·1)	4·03 (3·2-5)**	4·06 (3·1-5·1)**	6·13 (5·4-6·5)	6·25 (5·6-6·6)	5·41 (4·9-6·3)*	5·41 (5-6·1)*	N/A	N/A	N/A	N/A
HADS	12 (6-18)	10 (5-16)	14 (6·5-19)*	14 (7-20)*	13 (7-19·8)	9 (4·8-14·3)	14·5 (10·8-20·3)**	14 (4·8-20)*	5 (2-11)	1 (0-4)	4·5 (2·8-12·8)**	6·5 (2·8-15·8)**	4 (1-8·5)	4 (1·3-9·3)	NA	NA
SNOT-20	31 (19-43)	28 (16-39·3)	34 (23-45)**	35 (23·8-44·3)**	30 (17-48)	22·5 (10·8-42·3)	32 (20-49)*	34·5 (21·3-49·8)*	13 (5-21)	13 (5-19)	14·5 (8·25-29·8)	21 (9·5-31·3)	2 (0-8)	2 (0-7·8)	N/A	N/A
ESS	7 (4-10·5)	6 (3-10)	8 (4-11)**	8 (4-12)	8 (4-11)	8 (4-10·3)	7 (4-11)	7 (4-11)	5 (3-8)	5 (2-8)	5·5 (3-8)	6 (3·8-8·8)	5 (1-7)	4·5 (1-7)	N/A	N/A

Note: Data from each group are shown for the whole group (All), for the subgroup of participants with no history of GORD (NO-GORD), for the subgroup with a history/ diagnosis/ current treatment of GORD (ALL-GORD) and for those in whom symptoms of GORD were present at the time of assessment (ACTIVE-GORD). *= p<0·05 and ** =p<0·01 for the comparison between participants in the NO-GORD category as compared with ACTIVE-GORD or ALL-GORD, respectively. Data are shown as median (IQR) or mean (SD) depending on distribution. Abbreviations: ACQ = Asthma Quality Questionnaire; AQLQ = Asthma Quality of Life Questionnaire; HADS = Hospital Anxiety and Depression Scale; SNOT-20 = Sino-Nasal Outcomes Test; ESS = Epworth Sleepiness Score. N/A in group D implies that the questionnaire is not relevant for group or the sample size is too small for analysis.

RESULTS

Clinical characteristics and associations with GORD in the complete U-BIOPRED cohort

The prevalence of GORD was higher in both severe asthma groups (54% and 66%, respectively) when compared to mild/moderate asthmatic (18%) and healthy participants (5%) (Table 1). Similarly, active GORD was more prevalent in both severe asthma groups A and B (33% and 46%) (Table 1). Regardless of asthma severity, BMI in the asthmatics was raised at the nominal unadjusted significance level ($p<0{\cdot}05$) in the ALL-GORD and ACTIVE-GORD subgroups, as compared to NO-GORD, except in smoking severe asthmatics where BMI in the ACTIVE-GORD and NO-GORD subgroups was not different. Age, atopy, smoking, asthma exacerbation rates, spirometry, exhaled nitric oxide concentrations were not related to GORD; however, OCS use was more prevalent in GORD subgroups in the non-smoking severe asthmatics. Sputum neutrophil counts were significantly higher in mild/moderate asthmatics with active GORD when compared to those without GORD. Sputum eosinophil counts were lower in smoking severe asthmatics (ALL-GORD and ACTIVE-GORD). A number of patient reported outcomes were associated with GORD (figure 1, table 2): in smoking severe asthmatics, ACQ5, ACQ7, AQLQ, HADS and SNOT-20 scores were raised in patients with GORD, while in the non-smoking severe asthmatics GORD was associated with higher HADS, SNOT-20 and ESS. GORD was also associated with AQLQ and HADS in mild/moderate asthma.

Proteins associated with GORD

In order to remove any confounding effect of disease severity and because of low prevalence of GORD in healthy participants, proteomics data from mild/moderate asthmatics and healthy participants were excluded from the analysis. A total of 154 samples from severe asthmatics which passed QC were assessed (108 non-smokers and 46 smokers): of these, 90 had a diagnosis of GORD (ALL-GORD) and 55 also had active symptoms and/or were taking PPI (ACTIVE-GORD). This sub-cohort was slightly different from the complete U-BIOPRED cohort (Supplement Table 1): when compared with the NO-GORD subgroup, the GORD (All and Active) subgroups were not different in age, BMI, or lung function but more severe asthmatics were on OCS in the ALL-GORD when compared to NO-GORD subgroups.

Exacerbation frequency, ACQ5, ACQ7, AQLQ, HADS and SNOT-20 scores were higher in both ALL-GORD and ACTIVE-GORD sub-groups.

The primary comparison of sputum protein profiles between ACTIVE-GORD (n=55) and NO-GORD (n=64) subgroups identified 152 proteins detectable in ≥60% of participants, with 5 proteins being differentially abundant at p<0·05: Ig lambda-2 chain C regions was raised in ACTIVE-GORD, while alpha-1-antichymotrypsin, plasma protease C1 inhibitor, immunoglobulin lambda variable 1-47 and alpha-1-acid glycoprotein 1 were reduced (Table 3). These proteins were analysed by ULR together with another 6 proteins with significance at p≤0·1: lactotransferrin, lipocalin-1, serotransferrin, keratin type II cytoskeletal 6B, keratin type I cytoskeletal 10 and heat shock cognate 71 kDa protein (Table 3). Subsequent MLR, adjusted for smoking history and OCS treatment, identified four proteins associated with ACTIVE-GORD (Table 4): immunoglobulin lambda variable 1-47, plasma protease C1 inhibitor, lipocalin-1, and Ig lambda-2 chain C region. The first three proteins were retained in a further multiple regression analysis with backward selection (Figure 2).

A further Mann-Whitney U analysis, comparing sputum proteomes from the ALL-GORD and NO-GORD subgroups, yielded 10 differentially abundant proteins for ULR analysis (supplementary table 2, figure 1): 3 (immunoglobulin lambda variable 1-47, Alpha-1-antichymotrypsin and heat shock cognate 71 kDa protein) were related to the diagnosis of GORD (p≤0·05) and were associated with ALL-GORD when adjusted for smoking and OCS use (Table 4).

DISCUSSION

To our knowledge, this study provides the first evidence that suggests a biological effect of GORD within the lungs and the first evidence of that effect in severe asthma. A study by Parameswaran et al. suggested that lipid-laden macrophages (LLMs) are markers of oropharyngeal reflux,[19] although it did not report that numbers were increased in asthmatics. Similar to our study, there were no differences in inflammatory cell counts between participants with and without reflux. A further study by Gibeon et al. [20] of 21 severe and 17 mild/moderate asthmatics confirmed that LLMs were more frequent in patients with GORD but did not show any relationship with asthma severity even though the prevalence of GORD was three-fold higher in severe asthmatics, possibly because it was a small study and all participants were on treatment for reflux.

The observed low prevalence of GORD in mild/moderate asthmatics and healthy participants in our study, and the absence of its impact on clinical outcomes in these groups, makes it unlikely that GORD

Table 3. Comparison of protein abundance between ACTIVE-GORD and NO-GORD in the Severe asthmatics (Cohort A and B).

Mann-Whitney U comparison of proteins between ACTIVE-GORD and NO-GORD					
Protein name	**Uniprot ID**	**ACTIVE-GORD Protein Abundance (IU)**	**NO-GORD Protein Abundance (IU)**	**Z**	**p value**
Immunoglobulin lambda variable 1-47	P01700	26781 (15624-37298)	38693 (19555-48680)	-2·788	0·005
Alpha-1-antichymotrypsin	P01011	110332 86807-140619)	130569 (102641-165315)	-2·708	0·007
Alpha-1-acid glycoprotein 1	P02763	28638 (15646-38783)	39145 (28541-46295)	-2·510	0·012
Ig lambda-2 chain C regions	P0CG05	460018 (282052-564478)	397731 (238317-481500)	-2·191	0·028
Plasma protease C1 inhibitor	P05155	21081 (13898-37249)	30850 (15568-43890)	-2·111	0·035
Lactotransferrin	P02788	327432 (285341-387855)	385396 (293012-436902)	-1·903	0·057
Lipocalin-1	P31025	41899 (28461-88086)	39573 (20233-49739)	-1·812	0·070
Serotransferrin	P02787	184384 (153424-278884)	178905 (131771-227614)	-1·764	0·078
Keratin, type II cytoskeletal 6B	P04259	16914 (5055-41084)	36110 (7661-43661)	-1·743	0·081
Keratin, type I cytoskeletal 10	P13645	35580 (17494-48261)	44056 (25742-72086)	-1·695	0·090
Heat shock cognate 71 kDa protein	P11142	21238 (13041-32211)	23376 (16362-38133)	-1·690	0·091

Note: Protein abundance (IU = international units) is shown as Median (IQR).

Table 4. Proteins identified as best predictors of ACTIVE-GORD by multiple logistic regression analysis with backward selection.

Proteins Identified as predictors of ACTIVE-GORD *vs* NO-GORD using MLR with backward selection		
Protein name	**Uniprot ID**	**P value**
Immunoglobulin lambda variable 1-47	P01700	0·017
Plasma protease C1 inhibitor	P05155	0.043
Lipocalin-1	P31025	0·034
Proteins Identified as predictors of ALL-GORD *vs* NO-GORD on MLR with backward selection		
Immunoglobulin lambda variable 1-47	P01700	0·011
Alpha-1-antichymotrypsin	P01011	0·015
Heat shock cognate 71 kDa protein	P11142	0·02

has much impact on the lungs in milder disease or health. For this reason, detailed sputum proteomic analysis was limited to severe asthmatics in U-BIORED cohorts A and B, thereby removing disease severity as a bias. The 154 severe asthmatics whose sputum samples were assessed by mass spectrometry were representative of the wider cohort, with similar gender distribution, BMI, prevalence of GORD and ACTIVE-GORD, lung function, quality of life measures and sputum inflammatory cells. Asthma control, quality of life, and HADS scores were worse in ALL-GORD and ACTIVE-GORD subgroups. The higher SNOT-20 scores suggested a moderate effect of GORD on symptoms of rhinosinusitis. In contrast to the wider cohort, in the restricted group of severe asthmatics who provided sputum, those with GORD had more exacerbations and a higher proportion were on OCS than those without.

Eleven proteins were differentially abundant in severe asthmatics with active GORD compared to those without a GORD diagnosis at a significance level of $p<0{\cdot}1$; multiple logistic regression analysis with adjustment for smoking history and OCS use showed that three of these were associated with ACTIVE-GORD. While the concentrations of immunoglobulin lambda variable 1-47 and plasma protease C1 inhibitor were lower in ACTIVE-GORD when compared to participants with no history of GORD, the concentration of lipocalin-1 was higher. Analysis also identified three proteins associated with the diagnosis of GORD: immunoglobulin lambda variable 1-47, alpha-1-antichymotrypsin and the heat shock cognate 71 kDa protein, all of which were at lower concentrations in patients with GORD. The majority of proteins associated with GORD in this study have predominantly protective functions.[21] Some (heat shock cognate 71 kDa, alpha-1-acid glycoprotein, Ig light chains and serotransferrin) are transported into the epithelial lining fluid by transcytosis or by transudation from the circulation. Others, like lipocalin-1 and lactoferrin, are derived from mucosal glands and exert anti-microbial properties, while alpha-1-antichymotrypsin is an acute phase protein mainly produced in the liver with anti-inflammatory and proteolytic properties.[22]

Lactoferrin plays a role in innate immunity. It is produced by exocrine glands and is detected in all mucosal secretions. It is also stored in neutrophil secondary granules from which it has been shown to be released by allergen stimulation. Its concentrations are increased in bronchoalveolar lavage of asthmatics [23], so our finding of slightly lower concentrations in GORD suggests a different mechanism of regulation in the presence of reflux. In contrast, lipocalin-1, the archetypal member of the lipocalin superfamily, which also includes retinol binding proteins, apolipoprotein D and lactoglobulins[24], was increased in patients with GORD. In view of structural similarities to known antimicrobial peptides in the same superfamily and widespread distribution in the bronchial epithelium, its primary role is thought to be epithelial defence. We have previously shown reduced

levels in COPD but not in mild-moderate asthmatics.[17] The finding in the current study that levels are higher in severe asthmatics with GORD points to a distinct phenotypic feature of severe asthma which could underlie the increased risk of exacerbations found in our study. Lipocalin-1 sequesters siderophores produced by bacteria, thereby inhibiting bacterial growth through competition for iron reserves.[24] Lipocalins carry hydrophobic ligands such as lipids, steroids, hormones and other substances. When loaded with ligands, they induce regulatory T cells, leading to non-allergenic inflammation, whereas in empty state, they promote Th2 responses and inflammation.[25] Whether or not lipocalin-1 is bound to its ligands and whether this results in additional inflammation or tolerance in GORD requires further research.

Two subtypes of keratin were reduced in patients with GORD: keratin type-1 cytoskeletal 10 is widely distributed, while keratin type-2 cytoskeletal 6B is specific of distinct types of epithelia in the mouth and esophagus. As these are intracellular proteins, we speculate that their reduction reflects metaplasia where keratin isoforms could be altered, although there is at present no evidence of metaplastic epithelial changes in asthmatics with GORD. Complement C1 inhibitor, alpha-1-antichymotrypsin and alpha-1-acid glycoprotein are positive and serotransferrin is a negative acute phase marker. The finding that positive acute phase markers were lower in ACTIVE-GORD suggests that ACTIVE-GORD is associated with lower systemic inflammation. Elevated serotransferrin, the only negative acute phase marker, was elevated in ACTIVE-GORD, supporting this explanation.

Excess light chains synthesised during antibody synthesis are normally cleared rapidly by the kidney, while high levels of polyclonal Ig light chains are observed in a number of inflammatory conditions, including asthma. The conflicting patterns depending on the specific isoform seen in this study are possibly due to differential secretion of light chain isoforms from the endoplasmic reticulum during chronic inflammation or differential proteolytic cleavage events post-secretion. Binding of free light chains to neutrophils elicits IL-8 secretion *in vitro* and inhibits neutrophil apoptosis.[26,27]

The current study confirms the reported association between asthma and GORD,[7,28] with more than half severe asthmatics having a diagnosis of GORD and one third ACTIVE-GORD, and the smoking group having an even higher prevalence than non-smokers. As previously reported,[11,29] raised BMI was associated with GORD. Furthermore, GORD diagnosis and symptoms were associated with anxiety and depression, in both severe asthma groups and in mild/moderate asthmatics. Patients with severe asthma had more symptoms of rhinitis and sinusitis than mild/moderate asthmatics and healthy individuals and these were associated with GORD as previously reported.[9,10] While the association between SNOT-20 scores and GORD has been reported previously,[30,31] to our knowledge, this association has not, until this study, been extended to asthmatics.

This study has a number of limitations. Clinical assessment and management of GORD followed the standard guidelines[15] that do not recommend routine use of 24-hr pH/manometry study and, instead, recommend empirical medical treatment with a proton pump inhibitor. Given the severity of the disease and to minimise the number of invasive procedures in the already complex U-BIOPRED protocol, the GORD diagnosis was, therefore, based on history supplemented by therapy records at the time of recruitment. , It is possible that that some patients treated with PPI continued to have asymptomatic weakly- or non-acidic reflux which, the assessment of which would be useful. Furthermore, we did not assess separately participants on anti-reflux treatment, some of whom could have silent, weakly acid or non-acid reflux. Our analysis also treated as a single group, current and ex-smokers with significant smoking history. As this group had a higher prevalence of GORD in keeping with reported effects of smoking on GORD,[32] additional research is needed to assess the impact of cigarette smoking. The results for keratins and immunoglobulins will need confirmation because both are large families of highly homologous proteins with an increased probability of incorrect identification by MS that results in false positive results.

In summary, this study provides the first evidence which suggests that severe asthmatics with reflux have a distinct airways phenotype characterised by elevated anti-microbial proteins and reduced proteins that could be linked to systemic inflammatory responses and epithelial integrity, associated with poor asthma control, quality of life and additional co-morbidities. Further studies are required to confirm our findings, elucidate the roles of the differentially abundant proteins, and show whether the protein levels can be modulated by aggressive therapy of reflux.

Funding information

U-BIOPRED was supported through an Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115010, resources of which were composed of a financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution (www.imi.europa.eu). We would also like to acknowledge help from the IMI-funded eTRIKS project (EU Grant Code no. 115446).

Authors Contribution

KT contributed to patient recruitment, data collection, statistical analysis and data interpretation, and writing of the manuscript. JPRS contributed to data analysis and interpretation and writing of the manuscript. BLN contributed to data interpretation and writing of the manuscript. DB and JB

contributed to sample processing and analysis. BD designed and undertook the statistical analysis plan for the study. ATB contributed to statistical analysis and data interpretation. RD contributed to data interpretation and manuscript writing as well as overall supervision of the study. PJS (Paul J Skipp) supervised the overall proteomic analysis and contributed to data analysis and writing of the manuscript. SJW, RL, SJF, PSB, MC, BD, IH, NK, PM, MS, TS, TG, IP, ARS, IMA, DES, CA, PHH, PJS (Peter J Sterk) and KFC contributed as PI's with overall supervision in local centres where patients were recruited and contributed to writing of the manuscript.

Declaration of interests

KT, JB, JPRS, BLN, DB, PJS (PJ Skipp), AS, ATB, DS, MS, IP, PM, RL, SJF, SJW, TG and TS have no conflicts of interest to declare. RD reports personal fees from TEVA, grants and personal fees from Novartis, and personal fees and other support from Synairgen outside the submitted work. IMA reports grants from EU-IMI, during the conduct of the study and grants from MRC, BHF, Vectura and Dunhill Medical Trust; personal fees from Chiesi, GSK, Boehringer-Ingelheim, Vectura, and Astra Zeneca outside the submitted work. PB reports personal fees from AstraZeneca, personal fees from Boehringer-Ingelheim, during the conduct of the study. BD (B Dahlen) reports personal fees from Astra Zeneca, Teva, Novartis, outside the submitted work. CA reports grants from Innovative Medicines Initiative during the conduct of the study. IH reports personal fees from Astra Zeneca, GSK, Novartis, Sager Pharma, Chiesi, CSL Behring, Roche, Berlin Chemie, Boeringher-Ingelheim, Orion Pharma and MSD outside the submitted work. KFC reports personal fees from Advisory Board membership, grants for research, and personal fees from payments for lectures, outside the submitted work. MC reports grants from Innovative Medicines Initiative (IMI), during the conduct of the study. NK reports grants from IMI, during the conduct of the study. PHH reports part-time employment by GSK as Global Medical Expert. The institute of PJS has received a public-private grant for participating in this study from the Innovative Medicines Initiative (IMI) covered by the European Union (EU) and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

REFERENCES

1. Bel EH, Sousa A, Fleming L, et al. Diagnosis and definition of severe refractory asthma: an international consensus statement from the Innovative Medicine Initiative (IMI). *Thorax* 2011;66(10):910-7.
2. Field SK, Underwood M, Brant R, et al. Prevalence of gastroesophageal reflux symptoms in asthma. *Chest* 1996;109(2):316-22.
3. Harding SM, Guzzo MR, Richter JE. The prevalence of gastroesophageal reflux in asthma patients without reflux symptoms. *Am J Respir Crit Care Med* 2000;162(1):34-9.
4. Vincent D, Cohen-Jonathan AM, Leport J, et al. Gastro-oesophageal reflux prevalence and relationship with bronchial reactivity in asthma. *Eur Respir J* 1997;10(10):2255-9.
5. Sontag SJ, O'Connell S, Khandelwal S, et al. Most asthmatics have gastroesophageal reflux with or without bronchodilator therapy. *Gastroenterology* 1990;99(3):613-20.
6. Emilsson OI, Bengtsson A, Franklin KA, et al. Nocturnal gastroesophageal reflux, asthma and symptoms of obstructive sleep apnoea: a longitudinal, general population study. *Eur Respir J* 2013;41(6):1347-54.
7. ten Brinke A, Sterk PJ, Masclee AA, et al. Risk factors of frequent exacerbations in difficult-to-treat asthma. *Eur Respir J* 2005;26(5):812-8.
8. Shaw DE, Sousa AR, Fowler SJ, et al. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J* 2015;46(5):1308-21.
9. Moore WC, Meyers DA, Wenzel SE, et al. Identification of Asthma Phenotypes Using Cluster Analysis in the Severe Asthma Research Program. *American Journal of Respiratory and Critical Care Medicine* 2010;181(4):315-23.
10. Amelink M, de Nijs SB, de Groot JC, et al. Three phenotypes of adult-onset asthma. *Allergy* 2013;68(5):674-80.
11. Hampel H, Abraham NS, El-Serag HB. Meta-Analysis: Obesity and the Risk for Gastroesophageal Reflux Disease and Its Complications. *Ann Intern Med* 2005;143(3):199-211.
12. McQuaid KR, Laine L. Early heartburn relief with proton pump inhibitors: a systematic review and meta-analysis of clinical trials. *Clin Gastroenterol Hepatol* 2005;3(6):553-63.
13. Kiljander T, Rantanen T, Kellokumpu I, et al. Comparison of the effects of esomeprazole and fundoplication on airway responsiveness in patients with gastro-oesophageal reflux disease. *Clinical Respiratory Journal* 2012;7(3):218-7.
14. American Lung Association Asthma Clinical Research C, Mastronarde JG, Anthonisen NR, et al. Efficacy of esomeprazole for treatment of poorly controlled asthma. *N Engl J Med* 2009;360(15):1487-99.
15. Katz, P.O., Gerson, L.B. and Vela, M.F., 2013. Guidelines for the diagnosis and management of gastroesophageal reflux disease. *The American journal of gastroenterology*, *108*(3), p.308
16. Djukanovic R, Sterk PJ, Fahy JV, et al. Standardised methodology of sputum induction and processing. *Eur Respir J Suppl* 2002;37:1s-2s.
17. Nicholas BL, Skipp P, Barton S, et al. Identification of lipocalin and apolipoprotein A1 as biomarkers of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010;181(10):1049-60.
18. Schofield, J., Burg, D., Staykova, D., Corfield, J., Adcock, I., Chung, K.F., Sterk, P., Skipp, P. and Djukanovic, R., 2016. Proteome fingerprints define groups with distinct clinico-pathological phenotypes in the U-BIOPRED asthma study.
19. Parameswaran K, Anvari M, Efthimiadis A, et al. Lipid-laden macrophages in induced sputum are a marker of oropharyngeal reflux and possible gastric aspiration. *Eur Respir J* 2000;16(6):1119-22.
20. Gibeon D, Zhu J, Sogbesan A, et al. Lipid-laden bronchoalveolar macrophages in asthma and chronic cough. *Respir Med* 2014;108(1):71-7.

21. Joo NS, Evans IA, Cho HJ, et al. Proteomic analysis of pure human airway gland mucus reveals a large component of protective proteins. *PLoS One* 2015;10(2):e0116756.
22. Kalsheker N, Morley S, Morgan K. Gene regulation of the serine proteinase inhibitors alpha1-antitrypsin and alpha1-antichymotrypsin. *Biochem Soc Trans* 2002;30(2):93-8.
23. van de Graaf EA, Out TA, Kobesen A, et al. Lactoferrin and secretory IgA in the bronchoalveolar lavage fluid from patients with a stable asthma. *Lung* 1991;169(5):275-83.
24. Dittrich AM, Meyer HA, Hamelmann E. The role of lipocalins in airway disease. *Clin Exp Allergy* 2013;43(5):503-11.
25. Jensen-Jarolim E, Pacios LF, Bianchini R, et al. Structural similarities of human and mammalian lipocalins, and their function in innate immunity and allergy. *Allergy* 2016;71(3):286-94.
26. Braber S, Thio M, Blokhuis BR, et al. An association between neutrophils and immunoglobulin free light chains in the pathogenesis of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;185(8):817-24.
27. Cohen G, Rudnicki M, Horl WH. Uremic toxins modulate the spontaneous apoptotic cell death and essential functions of neutrophils. *Kidney Int Suppl* 2001;78:S48-52.
28. Wenzel SE, Busse WW. Severe asthma: Lessons from the Severe Asthma Research Program. *Journal of Allergy and Clinical Immunology* 2007;119(1):14-21.
29. Sutherland ER, Goleva E, King TS, et al. Cluster analysis of obesity and asthma phenotypes. *PLoS One* 2012;7(5):e36631.
30. Katle EJ, Hart H, Kjaergaard T, et al. Nose- and sinus-related quality of life and GERD. *Eur Arch Otorhinolaryngol* 2012;269(1):121-5.
31. Bohnhorst I, Jawad S, Lange B, et al. Prevalence of chronic rhinosinusitis in a population of patients with gastroesophageal reflux disease. *Am J Rhinol Allergy* 2015;29(3):e70-4.
32. Kahrilas PJ, Gupta RR. Mechanisms of acid reflux associated with cigarette smoking. *Gut* 1990;31(1):4-10.

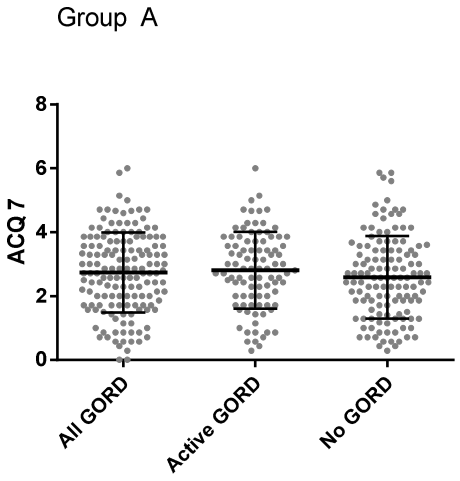

Figure 1a.

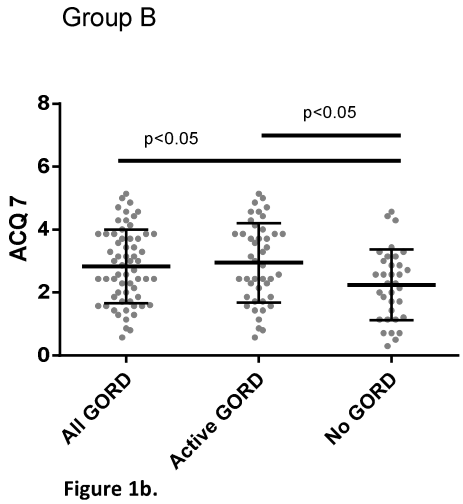

Figure 1b.

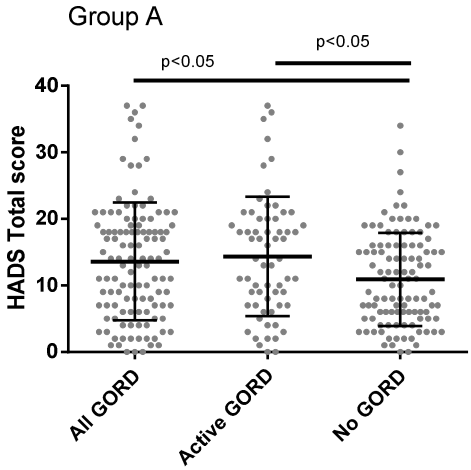

Figure 1c.

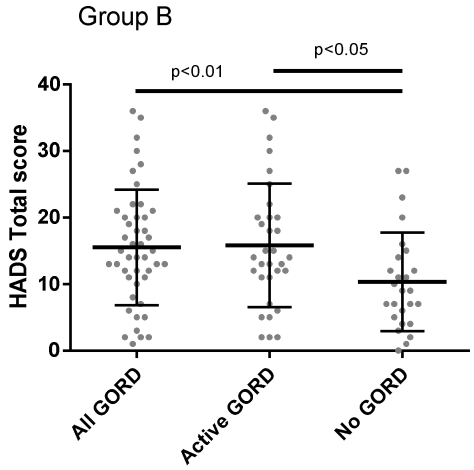

Figure 1d.

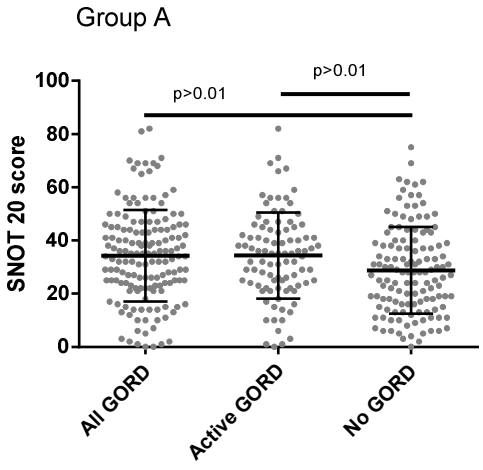

Figure 1e.

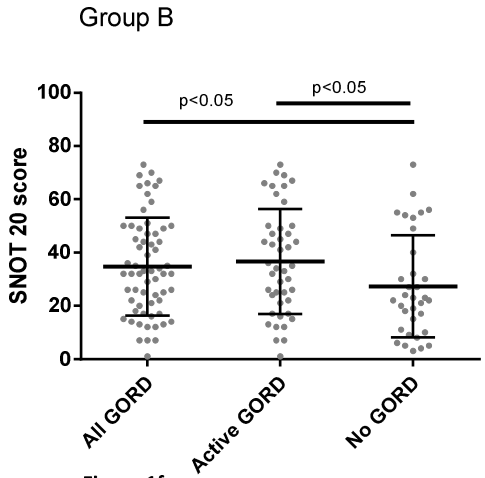

Figure 1f.

Figure 1. Patient reported outcomes for groups A and B: ACQ7 score (panels a-b**)**, HADS total score (c-d), SNOT 20 score (**e-f)**. For medians (IQR) please see table 2.

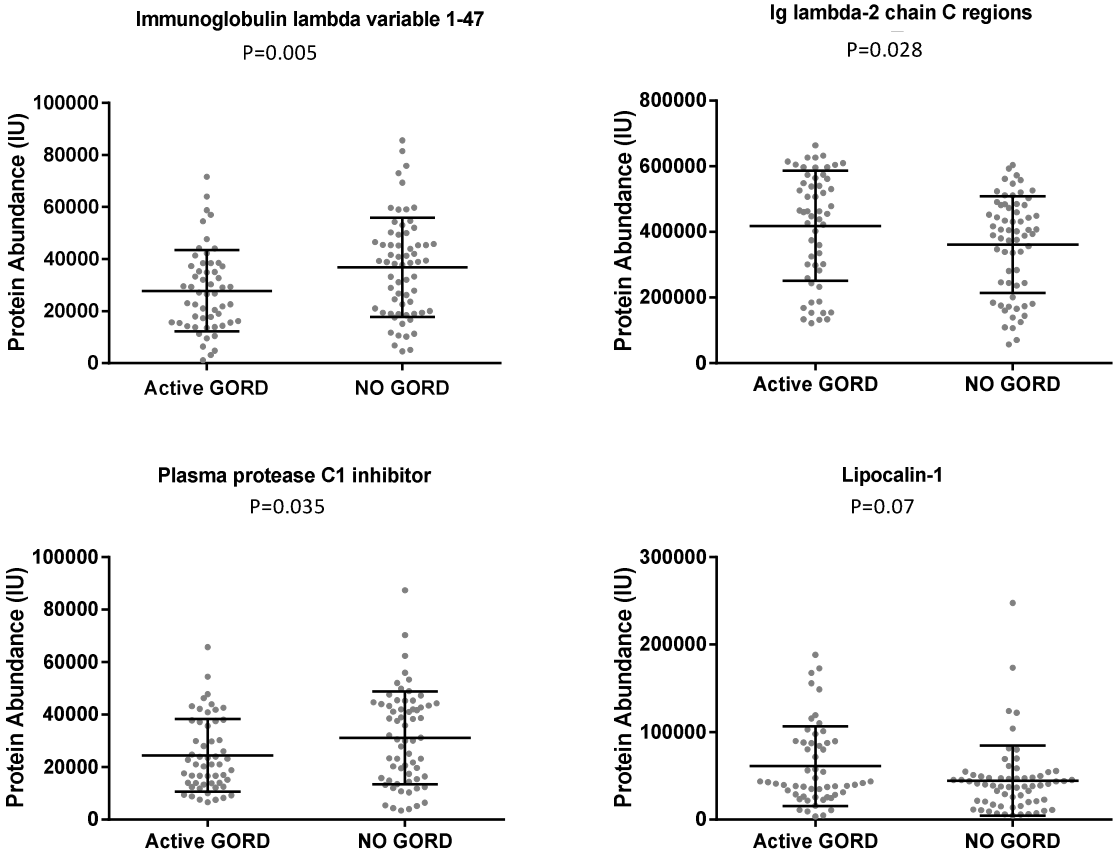

Figure 2. Proteins identified as best predictors of Active GORD. P values denote significance from initial Mann Whitney U tests.

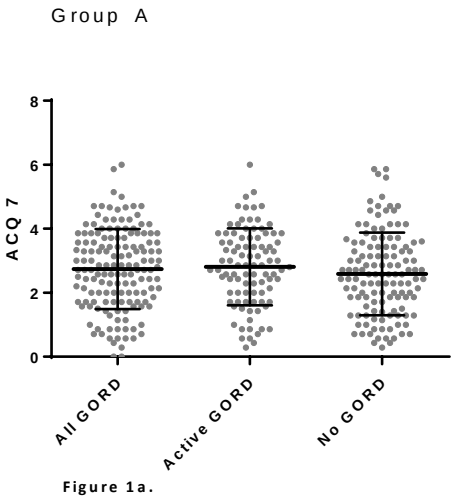

Figure 1a.

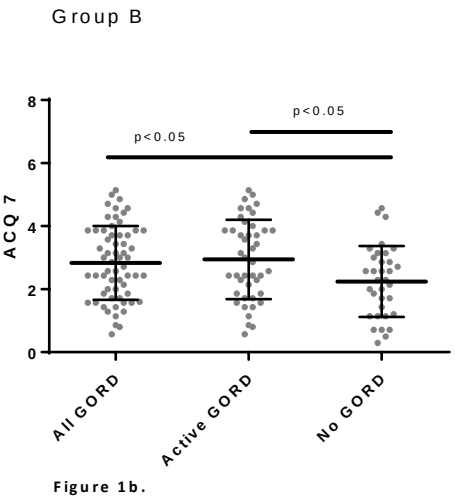

Figure 1b.

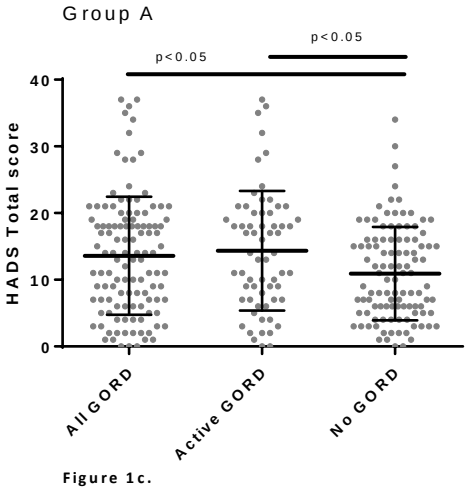

Figure 1c.

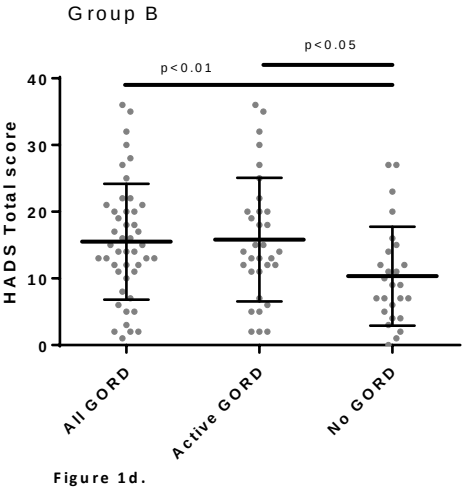

Figure 1d.

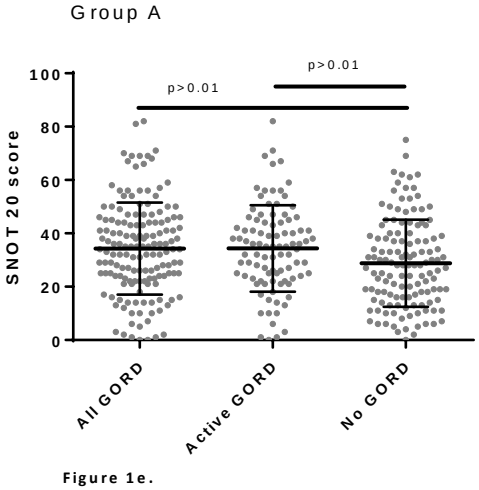

Figure 1e.

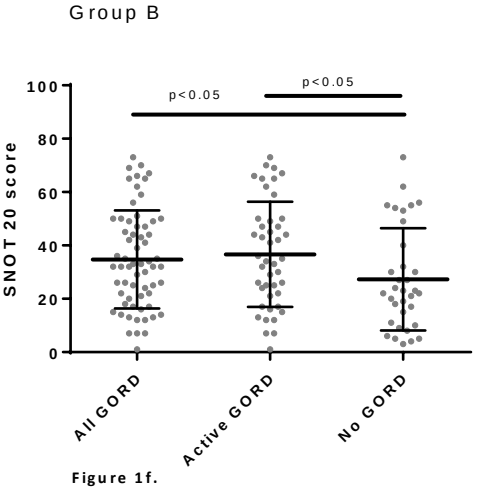

Figure 1f.

Figure 1. Patient reported outcomes for groups A and B: ACQ7 score (panels a-b**)**, HADS total score (c-d), SNOT 20 score (**e-f)**. For medians (IQR) please see table 2.

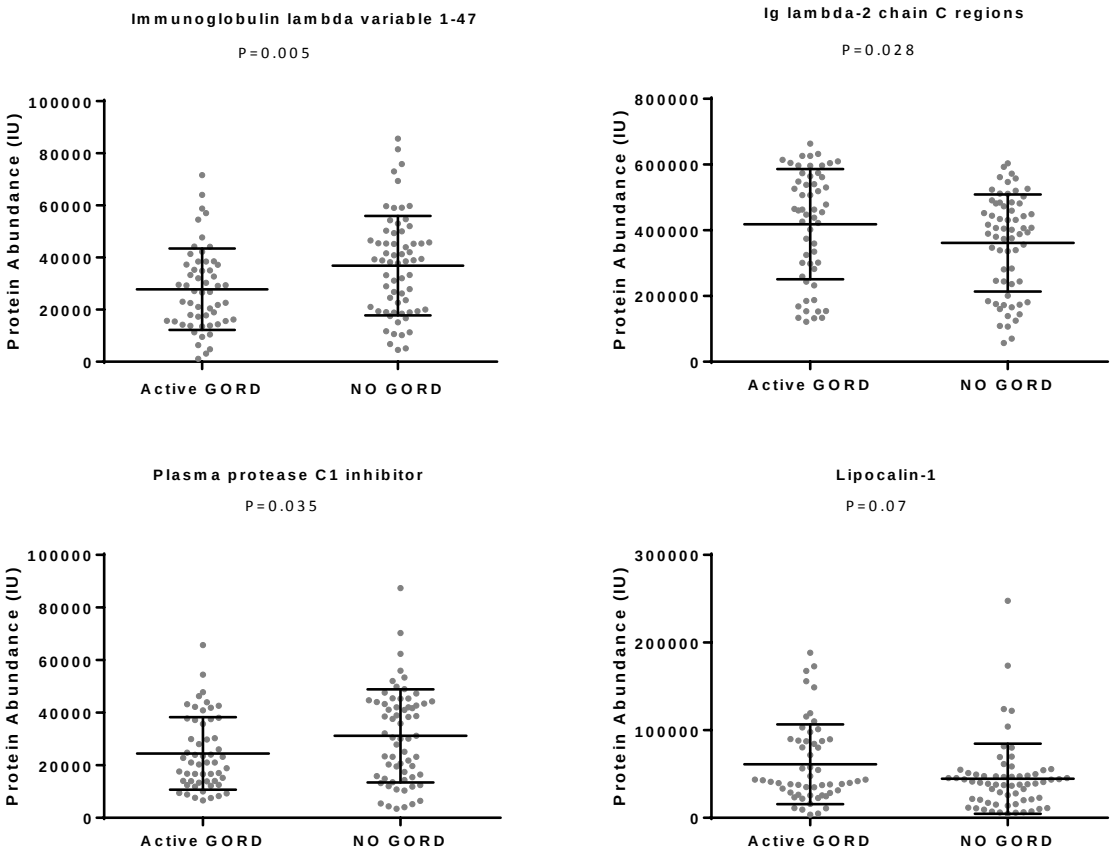

Figure 2. Proteins identified as best predictors of Active GORD. P values denote significance from initial Mann Whitney U tests.

Highlights

- 5 proteins different in severe asthmatics with GORD
- Ig lambda variable 1-47 was associated with active GORD in severe asthma
- lipocalin-1 was associated with active GORD in severe asthma
- plasma protease C1 inhibitor was associated with active GORD in severe asthma

Sputum proteomic signature of gastro-oesophageal reflux in patients with severe asthma - Supplementary material

Methods

Mass spectrometry

Peptide extracts were re-suspended in buffer A; 3% ACN, 0.1% Formic acid (v/v) and the concentration measured using a Direct Detect System (Millipore). An internal standard mixture of *E. coli* ClpB Hi3 standard (Waters), yeast enolase (ENO) and yeast alcohol dehydrogenase (ADH) was added to a final concentration of 250ng/µl sputum peptide in 20µl, 12.5 fmol/µl of ClpB, 12.5 fmol/µl of ENO, and 8.75 fmol/µl of ADH. Samples were analysed in duplicate via LC-IMS-MS^E on a Waters G2S high definition mass spectrometer coupled to a nanoAcquity UPLC system. 4 µl of peptide extract were injected onto a C18 BEH trapping column (Waters) and washed with buffer A for 5 min at 5 µl/min. Peptides were eluted using a 25 cm T3 HSS C18 analytical column (Waters), with a gradient of 3-50% ACN + 0.1% formic acid over 50 min. at a flow rate of 0.3µl/min. Eluted samples were sprayed directly into the mass spectrometer operating in MS^E mode. Data were acquired from 50 to 2000 *m/z* using alternate low and high collision energy (CE) scans. Low CE was 5V and elevated CE ramp from 15 to 40V. Ion mobility separation was implemented prior to fragmentation using a wave velocity of 650 m/s and wave height of 40V. The lock mass Glu-fibrinopeptide, $(M+2H)^{2+}$, m/z = 785.8426) was infused at a concentration of 100 fmol/µl with a flow rate of 250 nl/min and acquired every 60 sec.

Data curation and searching

Raw data were processed using a custom package (Regression tester) based upon executable files from ProteinLynx Global Server 3.0 (Waters). The optimal setting for peak detection across the dataset was determined using Threshold Inspector (Waters) and these thresholds were chosen: low energy = 100 counts; high energy = 30 and a total energy count threshold of 750. Database searches were performed using regression tester and searched against the Uniprot human reference database (20/11/2014) with added sequence information for internal standards. Quantity was estimated in absolute amounts using the Hi 3 method (1, 2). The ion accounting output files (3) were compiled and summary information generated from search log files using custom Python scripts. Information contained in ion accounting files were collated into a single .csv document using a custom Python script.

Data filtering and normalisation

Protein identifications collated from the ion accounting files were further quality filtered by allowing only identifications with the following criteria: identification in at least two separate samples (not including duplicates), a process that required at least three high quality unmodified peptides using the Hi3 method, and 2 peptides with at least 4 fragment ions for each protein. All other protein identities were removed. Proteins were ranked according to coverage across the samples and samples were ranked according to the order in which they were run. QC information was added for each sample (batch information, protein concentration, ion counts).

Inforsense software (ID Business Solutions, Guildford, UK) was applied to generated heat maps for the top 150 proteins using both 'top 3 peptide intensity sum' (a proxy for concentration) and peptide concentrations (expressed in fmol) on column calculated from internal standards. Sample-wise correlation plots were created using Inferno RDN (http://omics.pnl.gov/software/infernordn)(4). Heat maps and correlation plots were inspected for poor samples or injections; those with very low or no ID's and/or poor correlation were removed from the dataset.

Samples were analysed in duplicate and the intensity values from the injections was averaged. Replicate injections were inspected for consistency in quantitation, to enable this an average of the two injections 'top 3 peptide intensity sum' was calculated and a distance matrix was calculated by taking the Euclidian distance between the two injections as a function of the average of the injections. The resulting values were visualised in heat map, enabling rapid inspection of duplicates with high variance, which likely indicated a technical issue between injections (e.g. sprayer dropout, or failure to inject the correct volume). To uniformly remove suspect injections from the dataset we created the following universal rule: For samples with >2 fold between-injection difference in average intensity of proteins, the following rule (Rule 1) was applied: "report injection one intensity values for proteins, unless protein was only quantified in injection two, then include this value for increasing coverage".

While the above method was useful in identifying whole samples with poor repeatability between injections, there were cases where the concentrations of individual proteins were highly variable. To assess these cases, a log was created using a custom script, which highlighted those proteins where the ratio between injections was >1.5. Proteins with high frequency of poor measurement stability across all samples were processed according to 'Rule 2: "if the variation between injections is greater than 1.5 fold, take the quantity measured using injection one". The rationale behind taking the injection one values was that these are likely the cleanest: following on from an injection blank and extended equilibration, and less influenced by column carry over.

Mean values were derived from replicate sample injections except for those cases where rule 1 and rule 2 were applied, and those cases where the protein was quantified in only one sample; then the intensity value was taken for the single sample injection.

Differences in run-to-run intensity (loading) were adjusted by normalising each run to the sum of top 3 intensities of the proteins up to the point where the sample set reached 10% missing data (we refer to this as 'top-90 normalisation').

Supp. Table 1. Questionnaire and clinical characteristics of the severe asthma subset analysed for proteins predictive of GORD.

	Severe asthma (Group A non-smokers)				Severe asthma (Group B Smokers)				Severe asthma (Non-Smokers and smokers combined)			
	All in group	**NO GORD**	**All GORD**	**Active GORD**	**All in group**	**NO GORD**	**All GORD**	**Active GORD**	**All in group**	**NO GORD**	**All GORD**	**Active GORD**
N(%)	108 (70.12%)	48 (44.4%)	60 (55.55%)	34 (31.48%)	46 (29.87%)	16 (34.78%)	30 (65.22%)	21 (45.65%)	154 (100%)	64 (41.56%)	90 (58.4%)	55 (35.7%)
Age *(Yrs)*	55.5 (44.3-62)	54.5 (39.3-63.3)	56.5 (47.5-62)	57.5 (49.3-62)	55 (46.8-62.3)	55.5 (46.3-64.8)	55 (49.8-61.3)	55 (48-62)	55 (46-62)	55 (41.8-64)	55.5 (48.5-62)	55 (50-62)
Gender-M/F (%)	39/69 (36/64)	21/27	18/42	12/22	18/28 (39/61)	8/8	10/20	7/14	57/97 (37/63)	29/35	28/62	19/36
BMI	27.75 (23.9-33.3)	27.3 (23.7-31.9)	28.3 (24.2-33.9)	29.1 (23.1-33.8)	29 (25.4-33.5)	27.3 (25.4-32.5)	29.88 (25.5-36.7)	29.4 (25.1-37.6)	27.83 (24.6-33.3)	27.3 (24.1-32.1)	29.39 (25-34.1)	29.3 (24-34.4)
Smoking (PY)	0	0	0	0	20 (12.8-26.7)	21.25 (13.3-27.8)	26.25 (12.8-26.9)	15 (12.8-20.7)	0	0	0	0
On oral steroids - N (%)	42 (38.9%)	14 (33.3%)	28 (66.7%)	15 (18.3%)	21 (45.7%)	6 (28.6%)	15 (71.4%)	9 (24.3%)	63 (40.9%)	20 (13%)	43 (27.9%)*	24 (20.2%)
Exac. in last 12 months	2 (1-3)	2 (1-3)	2 (1-4)	2 (1-4)	2 (1-4)	2 (0-3.5)	2 (1-4.3)	2 (1-4)	2 (1-3)	2 (1-3)	2 (1-4)*	2 (1-4)
Exhaled NO	23.5 (15-41.3)	25.5 (14-44.8)	25.5 (14-44.8)	23.8 (19.3-38.3)	20 (10.5-44.8)	24.5 (10.3-33.8)	18.8 (10.4-71.8)	20 (10.8-55)	23 (14-41)	25.5 (12.3-36.4)	22.5 (14.5-45)	23 (14.5-41.8)
Atopy POS/NEG (% Pos)	69/21 (76.7%)	34/7 (37.8%)	35/14 (38.9%)	21/8 (23.3%)	24/12 (66.7%)	11/2 (30.6%)	13/10 (36.1%)	8/8 (22.2%)	93/33 (73.8%)	45/9 (35.7%)	48/24 * (38.1%)	29/16 * (23%)
FEV1 % *pred*	64.03 (21.4)	64.31 (22.2)	63.82 (20.9)	63.6 (20.8)	65.18 (17.7)	62.9 (20.1)	66.39 (16.5)	67.67 (18.6)	64.38 (20.3)	63.95 (21.6)	64.68 (19.5)	65.16 (19.9)
FVC *% pred*	87.1 (18.6)	88.62 (17.4)	85.9 (19.5)	85.1 (20.1)	89.9 (17.5)	91.25 (18.6)	89.18 (17.1)	89.6 (19.7)	87.94 (18.2)	89.28 (17.6)	86.9 (18.7)	86.82 (19.9)
Sputum-Macrophage *%*	26.98 (9.7-50.2)	26.12 (9.5-39.1)	27.52 (10.1-55.2)	37.31 (9.5-56.2)	34.08 (15.3-49.3)	30.25 (16.9-43.6)	34.84 (13.5-54.2)	34.91 (15.4-56.8)	27.83 (12.6-49.3)	27.52 (12.8-40.9)	30.04 (12.5-53.5)	36.77 (11.9-56)

	Severe asthma (Group A non-smokers)				Severe asthma (Group B Smokers)				Severe asthma (Non-Smokers and smokers combined)			
	All in group	**NO GORD**	**All GORD**	**Active GORD**	**All in group**	**NO GORD**	**All GORD**	**Active GORD**	**All in group**	**NO GORD**	**All GORD**	**Active GORD**
Sputum-Eosinophil %	2.59 (0.3-17.9)	2.08 (0.2-20.6)	3.04 (0.4-17.9)	3.47 (0.4-15.2)	3.55 (0.4-13.6)	4.89 (1.9-23.1)	1.89 (0.4-11.3)	1.79 (0.3-5.9) *	2.75 (0.4-16)	3.41 (0.4-21.1)	2.64 (0.4-14.9)	2.8 (0.4-8)
Sputum-Neutrophil %	56.69 (32.1-78.8)	62 (37.3-82)	54.47 (29.8-75.1)	51.5 (33.7-77.4)	55.55 (34.8-65.6)	55.52 (34.4-65.8)	56.52 (34.7-66.3)	59.44 (34.4-73.1)	55.1 (33.5-74.4)	58.17 (36.8-77.7)	52.48 (31.7-73.1)	52.67 (34.2-75.4)
ACQ5	2 (1-3)	2 (1-3)	3 (2-3)*	3 (2-4)*	2 (1-3)	2 (1-3)	2 (2-3)	3 (2-3.5)*	2 (1-3)	2 (1-3)	2 (2-3)**	3 (2-3.75)**
ACQ7	3 (2-3.3)	2 (1-3)	3 (2-4)	3 (2-4)	2 (2-4)	2 (1.3-3)	3 (2-4)*	3 (2-4)*	2 (2-3.5)	2 (1-3)	3 (2-4)*	3 (2-4)**
AQLQ-Total	5 (4-6)	5 (4-6)	4 (3-5)**	4 (4-5)*	4.5 (4-5.8))	5 (4-6)	4 (3.5-5)	4 (3-5)	5 (4-6)	5 (4-6)	4 (3-5)**	4 (4-5)*
AQLQ-Activity	4 (3-6)	5 (4-6)	4 (3-5)*	4 (3-5)	4 (4-6)	5 (4-6)	4 (3-5.5)	4 (3-6)	4 (4-6)	5 (4-6)	4 (3-5)**	4 (3-5)*
AQLQ-Symptom	5 (4-6)	5 (4-6)	4 (3-5.5)*	4 (3.3-5)*	4.5 (3.3-5)	5 (4-6)	4 (3-5)*	4 (3-5)	5 (4-6)	5 (4-6)	4 (3-5)**	4 (3-5)**
AQLQ-Emotional	5 (4-6)	5 (4-6)	4 (3-6)**	4 (3-6)*	5 (4-6)	5 (4-6)	5 (3.5-5.5)	5 (3-5)	5 (4-6)	5 (4-6)	5 (3-6)**	5 (3-6)**
AQLQ-Environment	5 (4-6)	5 (4-7)	5 (3-6)*	5 (4-6)	5 (3.25-6)	5 (4-7)	4 (3-5.5)	4 (3-6)	5 (4-6)	5 (4-7)	5 (3-6)**	5 (4-6)*
HADS	12.5 (6-18)	11 (6-16.5)	14 (6-18.3)	17 (7-20)	12 (7-17.3)	8 (4-11.3)	13 (10.3-20)**	14 (12-22.5)***	12 (6.8-18)	9.5 (6-14.8)	13.5 (7-19.3)*	15 (7.5-20)*
SNOT20	31 (23-43)	30 (16-38)	33.5 (24-47.5)	36.5 (24.8-46)	24 (12-42)	16.5 (5.3-25.8)	32 (18-47)**	35 (20-48.5)**	29 (18-43)	25.5 (14-37)	32 (24-47)**	36 (24-47)**
ESS	7.5 (5-10.8)	6 (3.5-10)	9 (6-11)	9 (5.5-12)	8.8 (4-11.3)	9 (5.5-10.8)	7 (3.8-12)	8 (4.3-11.8)	8 (4.8-11)	7 (4-10)	9 (5.5-11.5)	9 (5-12)

Note (Table 1): Data from each group are shown for the whole group (All), for the subgroup of participants with no history of GORD (NO GORD), for the subgroup with a history of GORD (ALL GORD) and for those in whom symptoms of GORD are present at the time of assessment (Active GORD). In the comparison of NO GORD and Active GORD or ALL GORD * represents $p<0.05$, ** represents $p<0.01$ and *** represents $p<0.001$. Atopy data excluded data points where the result of combined atopy was "uncertain". ACQ (Asthma control questionnaire), AQLQ (Asthma quality of life questionnaire), HADS (Hospital anxiety and depression score), SNOT (Sino-nasal outcome test), ESS (Epworth sleep score)

Table 2. Mann-Whitney U analysis of severe asthmatics (Cohort A and B) – All GORD *vs* NO GORD. Proteins with p≤0.1 were selected for ULR.

Mann-Whitney U test of proteins between All GORD and NO GORD					
Protein name	**Uniprot ID**	**All GORD Protein Abundance (IU)**	**NO GORD Protein Abundance (IU)**	**Z**	**p value**
Immunoglobulin lambda variable 1-47	P01700	26673 (15028-37586)	38693 (19555-48680)	-3.105	.002
Heat shock cognate 71 kDa protein	P11142	17474 (12671-31336)	23376 (16362-38133)	-2.317	.021
Ig lambda-2 chain C regions	P0CG05	452603 (255358-534754)	397731 (238317-481500)	-2.068	.039
Lactotransferrin (Lactoferrin)	P02788	328357 (259707-405139)	385396 (293012-436902)	-2.049	.040
Serotransferrin (Transferrin)	P02787	189065 (153260-273889)	178905 (131771-227614)	-1.991	.047
Alpha-1-antichymotrypsin	P01011	123534 (92929-146781)	130569 (102641-165315)	-1.870	.062
Immunoglobulin heavy variable 3-13	P01766	57006 (38913-89004)	45428 (34467-77853)	-1.855	.064
Ig gamma-1 chain C region	P01857	277390 (183825-341790)	238961 (181477-308449)	-1.763	.078
Pulmonary surfactant-associated protein A2	Q8IWL1	82788 (48836-148667)	69890 (39243-116216)	-1.639	.101
Alpha-1-acid glycoprotein 1	P02763	32817 (18534-45993)	39145 (28541-46295)	-1.635	.102

Univariate logistic regression analysis of proteins between All GORD and NO GORD

Proteins identified	Uniprot ID	p value	Odds Ratio	95% CI for Odds ratio	
Alpha-1-antichymotrypsin	P01011	.029	0.999991593	0.999984069	0.999999117
Heat shock cognate 71 kDa protein	P11142	.048	0.99997596	0.999952131	0.999999789
Ig lambda-2 chain C regions	P0CG05	.068	1.000001939	0.999999855	1.000004023
Serotransferrin	P02787	.072	1.000003671	0.99999967	1.000007673
Pulmonary surfactant-associated protein A2	Q8IWL1	.075	1.000004237	0.999999574	1.000008899
Lactotransferrin	P02788	.078	0.999997137	0.999993948	1.000000325
Immunoglobulin heavy variable 3-13	P01766	.115	1.000007162	0.999998251	1.000016072
Ig gamma-1 chain C region	P01857	.172	1.000002224	0.999999029	1.000005419
Alpha-1-acid glycoprotein 1	P02763	.459	0.999994979	0.999981694	1.000008264

Table 3. ULR analysis of proteins for All GORD *vs* NO GORD. Proteins with p≤0.05 were selected for MLR

Table 4. ULR analysis of proteins for Active GORD *vs* NO GORD. Proteins with p≤0.05 were selected for MLR

Univariate logistic regression analysis of proteins between Active GORD and NO GORD

Protein name	Uniprot ID	p value	Odds Ratio	95% CI for Odds ratio	
				Lower value	Upper value
Immunoglobulin lambda variable 1-47	P01700	0.008	0.999969752	0.999947482	0.999992023
Alpha-1-antichymotrypsin	P01011	0.009	0.999988138	0.999979281	0.999996995
Plasma protease C1 inhibitor	P05155	0.028	0.999973111	0.999949159	0.999997064
Lipocalin-1	P31025	0.046	1.000009386	1.000000171	1.000018601
Ig lambda-2 chain C regions	P0CG05	0.053	1.00000233	0.999999969	1.000004691
Keratin, type I cytoskeletal 10	P13645	0.097	0.999992128	0.999982823	1.000001432
Serotransferrin	P02787	0.124	1.000003501	0.999999042	1.00000796
Lactotransferrin	P02788	0.124	0.999997102	0.999993408	1.000000796
Keratin, type II cytoskeletal 6B	P04259	0.177	0.999987011	0.999968167	1.000005855
Heat shock cognate 71 kDa protein	P11142	0.181	0.999981678	0.999954847	1.000008509
Alpha-1-acid glycoprotein 1	P02763	0.26	0.999991029	0.999975415	1.000006643

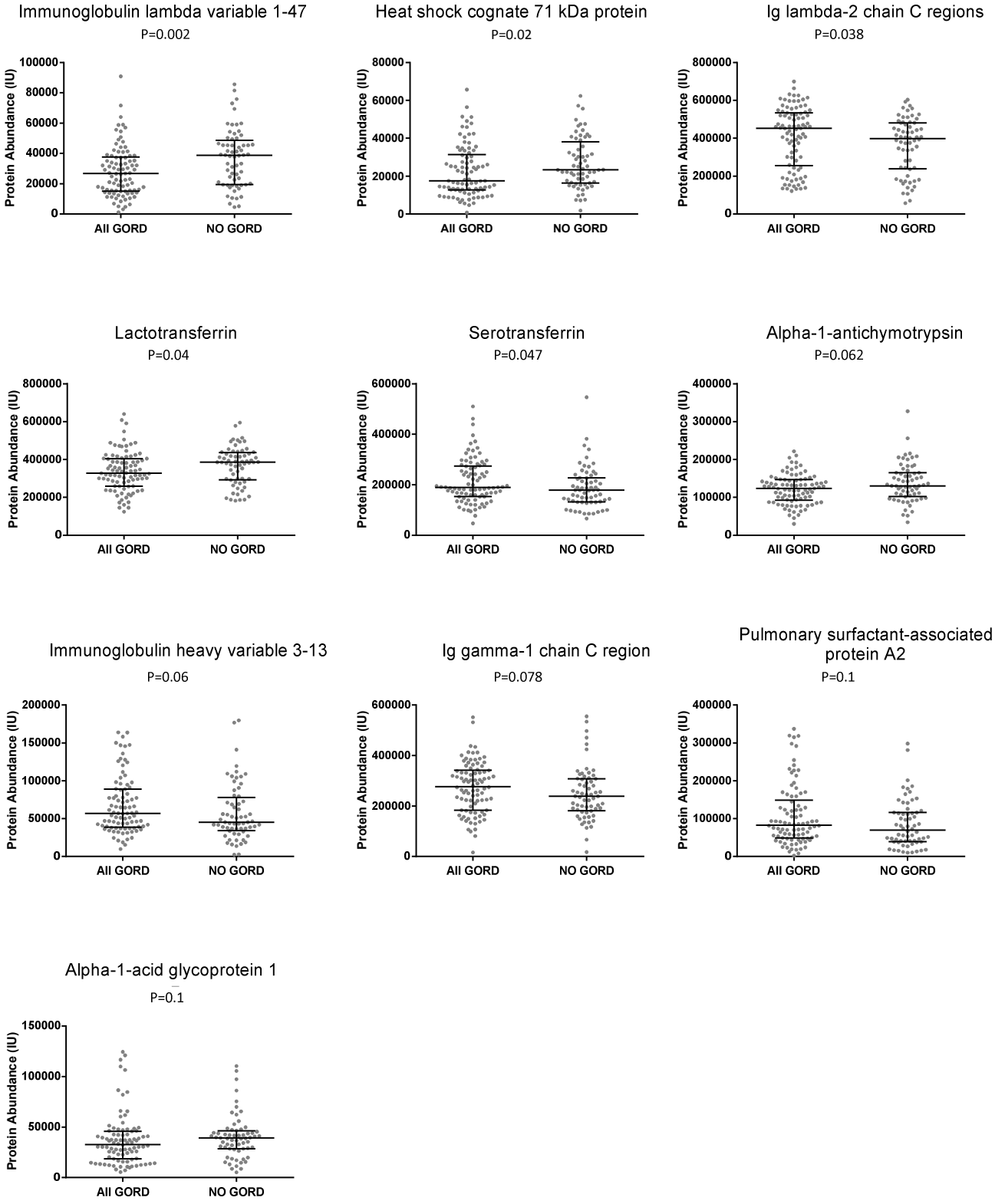

Figure 1. Proteins predictive of a diagnosis of GORD in severe asthma (All GORD) on Mann-Whitney U up to p≤0.1. *Immunoglobulin lambda variable 1-47* and *Heat shock cognate 71 kDa protein* were found to be the best predictors of a diagnosis of GORD after multiple logistic regression with backward selection.

REFERENCES:

Silva JC, Denny R, Dorschel C, Gorenstein MV, Li GZ, Richardson K, et al. Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale. Molecular & cellular proteomics : MCP. 2006;5(4):589-607.

Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, et al. Quantitative proteomic analysis by accurate mass retention time pairs. Analytical chemistry. 2005;77(7):2187-200.

Li GZ, Vissers JP, Silva JC, Golick D, Gorenstein MV, Geromanos SJ. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. Proteomics. 2009;9(6):1696-719.

Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp DG, 2nd, et al. DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics. 2008;24(13):1556-8.