A Novel Non-invasive Method Allowing for Discovery of Pathologically Relevant Proteins From Small Airways

Jörgen Östling
AstraZeneca Sweden: AstraZeneca AB
Marleen Van Geest
AstraZeneca Sweden: AstraZeneca AB
Henric K Olsson
AstraZeneca Sweden: AstraZeneca AB
Sven-Erik Dahlen
Karolinska Institute: Karolinska Institutet
Emilia Viklund
Inst Medicine; Gothenburg University
Per M. Gustafsson
University of Gothenburg Institute of Clinical Sciences: Goteborgs universitet Institutionen for kliniska vetenskaper
Ekaterina Mirgorodskaya
Gothenburg University
Anna-Carin Olin (anna-carin.olin@amm.gu.se)
University of Gothenburg https://orcid.org/0000-0001-6674-9020

Research

Keywords: Exhaled air, Non-volatiles, Breath, Small airways, Biomarker, Non-invasive, Asthma, Proteomics

Posted Date: December 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-78071/v3

Abstract

Background

There is a lack of early and precise biomarkers for personalized respiratory medicine. Breath contains an aerosol of droplet particles, which are formed from the epithelial lining fluid when the small airways close and re-open during inhalation succeeding a full expiration. These particles can be collected by impaction using the PExA® method (Particles in Exhaled Air), and are derived from an area of high clinical interest previously difficult to access, making them a potential source of biomarkers reflecting pathological processes in the small airways.

Research question

Our aim was to investigate if PExA method is useful for discovery of biomarkers that reflect pathology of small airways.

Methods and analysis

10 healthy controls and 20 subjects with asthma, of whom 10 with small airway involvement as indicated by a high lung clearance index (LCI ≥2.9 z-score), were examined in a cross-sectional design, using the PExA instrument. The samples were analysed with the SOMAscan proteomics platform (SomaLogic Inc).

Results

Two hundred-seven proteins were detected in up to 80% of the samples. Nine proteins showed differential abundance in subjects with asthma and high LCI as compared to healthy controls. Two of these were less abundant (ALDOA4, C4), and seven more abundant (FIGF, SERPINA1, CD93, CCL18, F10, IgM, IL1RAP). sRAGE levels were lower in ex-smokers (n=14) than in never smokers (n=16). Gene Ontology (GO) annotation database analyses revealed that the PEx proteome is enriched in extracellular proteins associated with extracellular exosome-vesicles and innate immunity.

Conclusion

The applied analytical method was reproducible and allowed identification of pathologically interesting proteins in PEx samples from asthmatic subjects with high LCI. The results suggest that PEx based proteomics is a novel and promising approach to study respiratory diseases with small airway involvement.

Key Messages

Key question: Can the PExA method identify individual protein profiles that reflect pathology of small airways, using the Somascan platform?

What is the bottom line? Two hundred-seven proteins were detected in up to 80% of the PEx samples, with a strong overrepresentation of proteins related to innate immune responses, including nine proteins that that discriminated subjects with asthma and high LCI as compared to healthy controls.

Why read on: The results support that PEx based proteomics is a novel and promising approach to study respiratory diseases with small airway involvement.

Introduction

There is a growing interest in the role of small airways (inner diameter < 2 mm) in asthma and other lung diseases [1]. In asthma, involvement of small airways is associated with more severe disease and loss of control [2–6], but has also been demonstrated in moderate and mild asthma [7]. Small airway involvement is also a recognised feature of chronic obstructive lung disease (COPD) [8], and in other severe lung-disease including viral bronchiolitis (as observed in e.g. COVID-19), lung-fibrosis and hypersensitivity pneumonitis.

In the small airways, surfactant plays a crucial role for airway patency and innate immune responses [9]. Surfactant is a complex mixture of proteins and lipids that keeps small airways open by reducing surface tension, but it also plays an important role in innate immunity by enhancing phagocytosis of inhaled pathogens and particulate matter by special surfactant proteins [10] and by modulating immune responses [11–14]. Given its crucial role for airway patency and host defence, knowledge of the protein and lipid composition of surfactant is surprisingly limited.

Although the small airways are a key compartment for the onset and progression of respiratory diseases such as COPD and lung-fibrosis [15], early detection of pathological processes in the small airways remains difficult mainly due its inaccessibility. One option for retrieving biological material from the small airways is through bronchoscopy with sampling of biopsies or bronchoalveolar lavage fluid (BALF), but this method is invasive and not suited for point-of-care situations or clinical trials. Non-invasive physiological measurements reflecting small airway function exist (e.g. inert-gas washout techniques and impulse oscillometry), but these methods do not provide the molecular information about pathways needed for the further development of precision medicine. In particular, the introduction of biologics, targeting specific molecular pathways, have highlighted the need for biomarkers that reflect disease endotypes, to enable patient stratification.

Particles in Exhaled Air (PExA) is a novel sampling method allowing non-invasive retrieval of biological material from the small airways. In short, the method is based on impaction of an aerosol consisting of ultrafine droplets of respiratory tract lining fluid (RTLF) that are formed and exhaled after a breathing manoeuvre that promote airway closure and reopening of the small airways [16]. PExA method has been thoroughly described by Larsson *et. al.* [17].

The molecular composition of PEx samples have been explored in previous studies, and 120 different proteins could be detected in PEx samples pooled from several individuals by LC/MS (21). The protein

composition of these samples showed up to 80% similarities to BALF.

Highly abundant proteins, like SP-A have been successfully quantified with low intra-individual variability in PEx samples from single individuals by ELISA [17, 18] and show good correlation to SP-A levels in BALF [19]. The small airway origin of the PExA sample is supported both by its composition, resembling (BALF) but not bronchial wash (BW)[19], and that no amylase is detected by LC/MS [20]. It is also indirectly supported by the 1000-10 000 fold increase in number of exhaled and sampled particles when using a breathing manoeuvre that promote airway closure and re-opening [16, 21].

In the present study we sought to evaluate whether the PEx samples convey information on pathophysiological processes useful in biomarker discovery. SOMAscan (SomaLogic Inc) was identified as a potentially suitable proteomics platform for the study. As PEx samples mainly originate from the small airways, we hypothesised differences in protein composition of PEx samples would be easiest to observe between healthy subjects and patients with increased lung clearance index (LCI), a standard measure of global ventilation inhomogeneity, i.e an indirect measure of small airway involvement that also is considered a sensitive indicator of early lung damage. Based on this reasoning we chose to study the protein composition of samples from asthmatic subjects with high LCI compared to that of asthmatic subjects and heathy controls with normal LCI.

Methods And Analysis

At first, we evaluated the performance of the Somascan platform and reproducibility. The second step was a clinical evaluation in a cross-sectional design, where the pathological relevance and the differences in protein-profiles in PEx of non-asthmatic subjects with that of subjects with asthma with- or without high LCI, were compared.

Subjects

Twenty subjects with asthma and ten healthy controls were included in the clinical evaluation. All were recruited from our earlier studies on asthma, or by an advertisement in a daily paper. To identify subjects with small airway involvement all subjects were screened with multiple breath nitrogen wash test (MBNW), giving an index of heterogeneity of ventilation (LCI). Asthma subjects were stratified into two groups, whereof one with normal LCI (zscore <2), herein referred to as A-nLCI (n=10), and one with high LCI (zscore ≥ 2.9), herein referred to as AhLCI (n=10). All subjects with asthma reported a physician diagnose of asthma and were taking asthma medication regularly. We also included a control group (non-asthma) that did not report respiratory symptoms nor were taking medication for respiratory disease and had normal LCI zscore (i.e. LCI<2), herein referred to as NA (n=10)[22].

Exclusion criteria were current smoking or smoking within the last 10 years or >10 pack-years, diagnosis of systemic inflammatory disease, cardiovascular disease or pregnancy. Demographic and clinical data

including LCI z-scores are presented in Table 1. All participants gave their written informed consent and the study was approved by the Ethical Committee at Gothenburg University in Sweden.

Table 1
Demographic and clinical characteristics of the three study-groups, including result from statistical tests

				p-values		
Parameter	**Non-asthma (NA)**	**Asthma with normal LCI (A-nLCI)**	**Asthma with high LCI (A-hLCI)**	**A-hLCI vs. NA**	**A-hLCI vs. A-nLCI**	**A-nLCI vs. NA**
Number	10	10	10			
Gender (Male/Female)	7/3	4/6	3/7	n.s	n.s	n.s
Age	48.9 (4.43) [28-66]	38.1 (4.1) [20-59]	54.6 (3.23) [38-68]	n.s	0.0040	n.s
Age at onset of asthma, yrs	-	17.4 (5.16) [5-55]	24.3 (7.38) [2.0-60]	-	n.s	-
BMI	23.89 (0.77) [19.26-27.16]	24.24 (0.8) [21.15-28.34]	25.93 (0.97) [21.47-31.18]	n.s	n.s	n.s
Allergy y/n	3/7	7/3	6/4	n.s	n.s	n.s
Ex smoker y/n	3/7	4/6	7/3	n.s	n.s	n.s
FEV1 (% pred)	100.9 (2.9) [88-117]	93.6 (4.63) [79-123]	71.2 (5.31) [39-91]	0.0003	0.0051	n.s
FEV1/FVC (%)	75.13 (7.67) [7.77-93.8]	79.2 (2.11) [71-88]	62.69 (3.73) [35-71]	0.0024	0.0001	n.s
Reversibility (%)	2.4 (1.66) [-5 - 9]	7.5 (2.31) [1–21]	14.7 (2.68) [6–28]	0.0006	0.0137	n.s
LCI z-score	0.89 (0.12) [0.5-1.7]	1.04 (0.17) [0-1.8]	5.07 (0.53) [2.9-8.1]	0.0001	0.0001	n.s
S-Cond VT, z-score	-0.57 (0.42) [-3-1.7]	-0.13 (0.4) [-1.5-1.4]	3.77 (0.49) [0.7-5.8]	0.0002	0.0005	n.s
S-Acin VT, z-score	0.59 (0.18) [0-1.5]	0.48 (0.21) [0-1.8]	2.76 (0.81) [0-8.6]	0.0077	0.0065	n.s
GINA step	-	2.2 (0.29) [1–4]	2.9 (0.41) [1–4]	-	n.s	-
ACQ, mean (1-6)	-	0.82 (0.3) [0-3.17]	1.13 (0.24) [0-2.17]	-	n.s	-

Data are presented as, means with standard error given in parenthesis and range given in brackets. Incomplete data is indicated by n numbers given in parenthesis. Kruskal Wallis and Chi Square statistical tests were used for analysing the differences between continuous and categorical data, respectively. "n.s" indicate statistical test with p-value below 0.05. Dash (-) indicate "not applicable".

				p-values		
B-neutrophils (%)	3.01 (0.33) [1.5-4.4]	3.68 (0.35) [2.2-5.4]	4.53 (0.52) [2.5-7.3]	0.0493	n.s	n.s
B-eosinophils(%)	0.15 (0.04) [0.06-0.5]	0.27 (0.04) [0.1-0.6]	0.3 (0.07) [0.1-0.6]	0.0287	n.s	0.0156
FENO, ppb	17.7 (1.93) [8–24]	41.2 (8.72) [6-86]	41.7 (11.74) [11-113]	n.s	n.s	n.s
hsCRP	0.51 (0.08) [0.23-1.1] (n=9)	0.432 (0.107) [0.14-1.2]	2.35 (0.62) [0.45-5.40]	0.0054	0.0031	n.s
Average mass pg/particle	0.23 (0.01) [0.2-0.27]	0.22 (0.01) [0.18-0.3]	0.22 (0.01) [0.17-0.29]	n.s	n.s	n.s

Data are presented as, means with standard error given in parenthesis and range given in brackets. Incomplete data is indicated by n numbers given in parenthesis. Kruskal Wallis and Chi Square statistical tests were used for analysing the differences between continuous and categorical data, respectively. "n.s" indicate statistical test with p-value below 0.05. Dash (-) indicate "not applicable".

Clinical characterization

Spirometry was performed according to ERS guidelines, using Spirare spirometer (Spirare, Stockholm, Sweden) Forced vital capacity (FVC) and forced expired volume in one second were expressed as a percentage of the reference value (FEV1% pred) derived from the ECCS/ERS reference equations [23].

Multiple Breath Nitrogen Wash-out tests were performed using the Exhalyzer® D device (Eco Medics AG, Duernten, Switzerland) and software (Spiroware 3.1) in accordance with current guidelines [24]. Z-scores were calculated as described by Kjellberg *et. al.* [25].

Fraction of exhaled nitric oxide (FENO) was measured once by a NIOX Mino (Aerocrine AB, Stockholm, Sweden) before spirometry following the ATS-ERS guidelines [26], except for only performing one exhalation.

A skin-prick test (SPT) to common allergens in Sweden was performed with positive result defined as a wheal diameter ≥3 mm and negative control <3 mm. Atopy was defined as the occurrence of at least one positive SPT wheal.

Serum samples were analysed for hsCRP and differential cell counts, using standard clinical methods.

All subjects filled out a questionnaire on medical history, smoking habits, symptoms and medication and subjects with asthma filled out Asthma Control Questionnaire, ACQ, reflecting asthma control over the

last week [27]. The use of medication was translated to GINA step for each subject according to GINA guidelines 2016.

PEx sample collection

The PExA method and PExA 1.0 instrument was used to collect PEx samples (described in supplement). For assessment of reproducibility, 120 ng of PEx was collected from each subject and for all other samples at least 240 ng of PEx was collected, involving two consecutive sampling sessions with a short break in between. After collection the sample holder was transferred to a clean-air room and the substrate was excised with a scalpel from the sample holder and placed in Millipore Ultrafree-MC LH Centrifugal Filter insert (FC30LH25) and stored at -80°C for further analysis. True blank samples were generated by applying the same procedure as for real samples but without a human breathing into the PExA instrument.

SOMAscan analysis and data processing

SOMAscan is a proprietary highly multiplexed, sensitive proteomic platform (SomaLogic Inc., Boulder, USA). As the SOMAscan platform developed during the study period two different versions was used; i) SOMAscan 1.1K was used for the assessment of SOMAscan performance with PEx samples and SOMAscan 1.3K for the other experiments. Platform and sample preparation is described in supplement.

Intra-run normalization and inter-run calibration were performed by SomaLogic according to their SOMAscan assay GLP data quality-control procedures. Data from SomaLogic was reported in relative fluorescent units (RFU) after hybridization control normalization which remove individual sample variance on the basis of signalling differences between scans (herein referred to as RFU values). Data from all samples passed quality-control criteria and were considered eligible for further analysis. Limit of detection (LOD) was calculated as 3 times the standard deviation from the mean RFU signal measured from 3 blank samples. Proteins with RFU values below LOD were not considered for further analyses. To account for systematic differences due to possible variability in final PEx concentration, the set of detected proteins were subjected to group median based normalization and log2 transformation before statistical analysis was performed. Mean and median values for establishment of LOD and normalization, respectively were calculated based on RFU values in all samples.

Gene Ontology enrichment analysis

To improve our understanding of the origin and functions of the proteins seen in PEx samples, a protein annotation enrichment analysis was performed, using the publicly available "Gene Ontology enrichment analysis and visualization tool GOrilla [28], matching a list of 199 uniquely mapped PEx proteins to either

the Cellular Component (CC) or the Biological Process (BP) GO sub-domain (database updated on Feb 15, 2020). A list of 1291 uniquely mapped SOMAscan protein identities was used as reference/background.

Statistical analysis

Significance level for the Gene Ontology enrichment analysis was calculated using the right-tailed Fisher exact test, provided by the GOrilla web-based service [28]. Result from GO annotation enrichment analysis were considered significant at a Benjamini–Hochberg corrected p-value below 0.05. PEx protein composition was compared to that of BAL and enrichment factor was calculated by Fisher Exact test.

SOMAscan data were mainly analysed using Qlucore Omics Explorer 3.6 software (Qlucore, Lund, Sweden), and was log2 transformed.. One-way analysis of variance (ANOVA) tests were used to determine intra-individual differences in the reproducibility experiment. General linear model statistics with each variable normalized to mean 0 and variance of 1 and adjustment for imbalance in age and BMI, was used to test differences between the NA, A-nLCI and A-hLCI groups. Benjamini-Hochberg multiple correction was used to control for rate of false-positive results (herein referred to as q-value). Statistical analysis of clinical and demographic variables was performed with Kruskal-Wallis or Chi-square tests using Spotfire 7.0.2 software (TIBCO Spotfire).

Group comparisons of SOMAscan data were considered hypothesis free and proteins with p value below 0.05 and a q-value below 0.2 was considered to be of interest in this explorative study.

Results

1. Assessment of SOMAscan assay performance for PEx samples

SOMAscan technical variability was evaluated by repeated measurements of a pooled PEx sample (1 µg PEx per ml) 5 times on the SOMAscan 1.1K platform. The mean CV value was 10% looking at a set of 174 proteins detected in all five samples, and below 20% for 156 of the 174 detected proteins (Figure 1). Intra-individual repeatability related to the PEx sampling procedure and the SOMAscan 1.1K platform combined, was evaluated by repeat measurements of three consecutive 120 ng PEx samples collected from 6 subjects with asthma. The intra individual CV values ranged from 6.1 to 24.8% with a mean of 13.8%, looking at a set of 114 proteins detected in all 18 samples. To assess if the observed intra-individual variability is low enough for the method to be useful for biomarker discovery, we analysed to what degree it was possible to separate the 6 subjects from each other, solely based on the proteomics data. Defining each of the 6 triplicate samples as groups, the between groups ANOVA test revealed 102 proteins with statistically significant differences between at least two of the group means ($q<0.05$). Filtering the list of protein variables further down to a q-value cut-off of $5.5 \cdot 10^{-5}$ yielded 42 proteins that completely separated all 6 subjects from each other, as judged by visual inspection of a Principal Component Analysis (PCA) plot (Figure 2).

2. Assessment of pathological relevance of proteins detected in PEx samples

Of the 1310 proteins represented on the 1.3K SOMAscan panel, 134 proteins showed RFU values larger than LOD in the complete set of 30 samples (2 µg PEx/ml). To increase chance of finding differentially abundant proteins a set of 207 SOMAscan protein ID's, detected over LOD in 80% of the 30 samples were used for various comparative data analyses (Table S1A).

2.1 Comparison of the protein composition of PEx with that of BALF by enrichment analysis

Of 207 proteins detected with the SOMAscan 1.3K platform, 81 (41%) have previously been detected in BALF [29]. Using 1323 uniquely mapped SOMAscan protein identities as reference/background gave at hand that the 207 proteins detected in PEx samples are enriched 5.9 times with the proteins previously detected in supernatant from BALF samples ($p<0.0001$).

2.2 Gene Ontology (GO) enrichment analysis

Gene Ontology enrichment analysis of 199 uniquely mapped PEx/SOMAscan protein ID's (Figure 3) revealed an over-representation of several Cellular Components (CC) GO terms, for example; "extra cellular exosome" (enrichment factor (EF)=1.79, q=6.30E-11), "blood microparticle" (EF=3.43, q=8.28E-10) and "platelet alpha granule lumen" (EF=3.15, q=1.78E-04) (Table 2A). Biological Process (BP) GO domain analysis revealed an over-representation of BP terms, for example; "regulation of complement activation" (EF=4.4, q=5.17E-08), "platelet degranulation" (EF=2.8, q=2.88E-04), "regulation of coagulation" (EF=2.6, q=2.72E-02), "acute inflammatory response" (EF=3.21, q=8.08E-03), and "neutrophil activation involved in immune response" (EF=1.69, q=2.9E-02) (Table 2B).

Table 2A

Gene ontology annotation enrichment analysis. Cellular Component sub-domain (CC)

Description (term)	Enrichment factor	Input and Output N,B,n,b	p-value	FDR q-value	GO term
extracellular exosome	1.79	1291,382,194,103	6.80E-14	6.30E-11	GO:0070062
extracellular vesicle	1.78	1291,384,194,103	1.03E-13	3.19E-11	GO:1903561
extracellular region	1.39	1291,566,194,118	1.92E-07	2.97E-05	GO:0005576
blood microparticle	3.43	1291,62,194,32	3.57E-12	8.28E-10	GO:0072562
platelet alpha granule lumen	3.15	1291,38,194,18	1.53E-06	1.78E-04	GO:0031093
extracellular matrix	1.82	1291,183,194,50	2.12E-06	2.18E-04	GO:0031012
collagen-containing extracellular matrix	1.91	1291,150,194,43	3.35E-06	3.11E-04	GO:0062023
cytoplasmic vesicle lumen	1.92	1291,135,194,39	8.75E-06	6.76E-04	GO:0060205
endoplasmic reticulum lumen	1.97	1291,91,194,27	1.61E-04	8.27E-03	GO:0005788
endoplasmic reticulum part	1.71	1291,132,194,34	4.70E-04	2.18E-02	GO:0044432

Table 2B

Gene ontology annotation enrichment analysis. Biological Process sub-domain (BP)

Description (term)	Enrichment factor	Input and Output N,B,n,b	p-value	FDR q-value	GO term
regulation of extracellular matrix constituent secretion	6.65	1291,5,194,5	7.33E-05	2.23E-02	GO:0003330
exocytosis	1.80	1291,185,194,50	3.05E-06	1.54E-03	GO:0006887
vesicle-mediated transport	1.74	1291,276,194,72	3.05E-08	2.78E-05	GO:0016192
secretion by cell	1.63	1291,224,194,55	2.27E-05	8.63E-03	GO:0032940
regulation of protein activation cascade	4.50	1291,34,194,23	2.28E-12	2.08E-08	GO:2000257
regulation of complement activation	4.44	1291,33,194,22	1.13E-11	5.17E-08	GO:0030449
complement activation, alternative pathway	4.44	1291,12,194,8	6.55E-05	2.14E-02	GO:0006957
complement activation, classical pathway	3.90	1291,29,194,17	5.25E-08	4.36E-05	GO:0006958
regulation of humoral immune response	3.64	1291,42,194,23	1.17E-09	1.77E-06	GO:0002920
innate immune response	1.81	1291,114,194,31	3.04E-04	6.03E-02	GO:0045087
platelet degranulation	2.80	1291,57,194,24	3.78E-07	2.88E-04	GO:0002576
fibrinolysis	3.52	1291,17,194,9	2.67E-04	5.41E-02	GO:0042730
negative regulation of coagulation	3.21	1291,29,194,14	1.86E-05	7.37E-03	GO:0050819
regulation of coagulation	2.60	1291,41,194,16	1.19E-04	2.72E-02	GO:0050818
regulation of haemostasis	2.60	1291,41,194,16	1.19E-04	2.86E-02	GO:1900046
acute-phase response	3.33	1291,24,194,12	4.93E-05	1.67E-02	GO:0006953
acute inflammatory response	3.21	1291,29,194,14	1.86E-05	8.08E-03	GO:0002526

neutrophil activation involved in immune response	1.69	1291,126,194,32	9.28E-04	1.69E-01	GO:0002283
regulation of response to external stimulus	1.48	1291,287,194,64	1.14E-04	2.90E-02	GO:0032101
defence response	1.40	1291,300,194,63	8.92E-04	1.70E-01	GO:0006952

Table 2 display result from Gene Ontology enrichment analysis using the publicly available "Gene Ontology enrichment analysis and visualization tool" (GOrilla) [28]. A list of 199 uniquely mapped PEx proteins detected with SOMAscan 1.3K were searched against the Cellular Component sub-domain database (2A) and the Biological Process sub-domain database (2B). A list of 1291 uniquely mapped SOMAscan 1.3K protein identities was used as reference/background. Enrichment factor was calculated as (b/n) / (B/N), where n - is the total number of PEx protein ID's, identified by SOMAscan and used as input, b - is the number of PEx /SOMAscan protein ID's associated with the GO term. p values for enrichment analysis were computed according to the mHG or HG model. FDR q value is the p value corrected for multiple testing using the Benjamini and Hochberg (1995) method

2.3 Differential abundance analysis, asthma vs. non-asthma

To identify confounding demographic factors we investigated the impact of gender, BMI and age, and found a clear effect of age and to some extent of BMI, independent of disease status. The relative abundance of each of the 207 detected proteins were then compared between various pairwise combinations of the A-hLCI, (n=10), A-nLCI (n=10) and NA (n=10) groups. Adjusting for imbalance in age, 9 proteins were found to be differentially abundant in A-hLCI as compared to the NA group, whereof 2 were less abundant (ALDOA4, C4) and 7 more abundant in A-hLCI (FIGF, SERPINA1, CD93, CCL18, F10, IgM, IL1RAP) (Table S1), exemplified in Figure 4. Reviewing the scientific literature revealed that all of the 9 differentially abundant proteins are known to play role in immune response and respiratory disease (Table S2).

2.4 Differential abundance analysis, ex-smokers vs. never smokers

To explore effect of smoking in a post-hoc analysis, the 207 SOMAscan/PEx protein data set was screened for proteins showing differential abundance in ex-smokers (n=14) vs. never smokers (n=16), see Table S3. Only one protein, sRAGE (soluble Receptor for Advanced Glycation End products), a pattern-recognition receptor involved in host response to injury, infection and inflammation fulfilled the significance criteria after adjusting for age and BMI, with decreased abundance in ex-smokers as compared to never smokers (Figure 4). By contrast, sRAGE did not show any clear difference between any of the asthma groups and healthy controls.

Discussion

Exhalation after breath-holding at residual volume give rise to release of high numbers of tiny droplets/particles formed from the respiratory tract lining fluid (RTLF) covering the small airways. Some

of these particles are small enough to follow the airstream of the exhalation and can be collected by impaction technology (PExA). Due to the small size of the particles and the specific origin, the total amount of RTLF that can be collected in this way is minute. In the present study we addressed the feasibility of proteomic profiling of PEx samples and could demonstrate that the SOMAscan proteomics platform is sensitive enough to detect and accurately quantify over 150 proteins in PEx samples from single individuals. Analysis of three consecutive samples indicated that intra-individual variability is substantially smaller than the inter-individual variability.

Moreover, protein enrichment analysis showed that protein composition of the PEx matrix resembles that of BALF supernatant to a large extent. This finding provide further confidence and confirms previous findings that PEx sample originate from small airways [19, 20] and hold the potential to be developed into a non-invasive substitute for bronchoscopy based diagnosis.

Protein enrichment analysis revealing that the PEx matrix is enriched in extracellular proteins associated with “exosome” (Figure 3, Table 2A), is of particular interest due to the emerging role of the exosomes as mediators of biomarkers for several chronic lung diseases [30, 31]. In addition, PEx proteome seem highly relevant for studies on the role of innate immune response in development of respiratory diseases and host defence.

To explore the pathological relevance of the PEx proteome in studies of respiratory disease we analysed PEx/SOMAscan data from 20 asthma patients and 10 healthy control subjects. Despite the low number of subjects, we found several highly interesting proteins to be differently abundant in samples from subjects with asthma compared to the non-asthma group. Alfa-1-antitrypsin (SERPIN1A) and IL1RAP were elevated only in asthma patients with high LCI, as opposed to IgM, CD93 and CCL18 which were elevated also in asthma patients without small airway dysfunction (Figure 4A). The two different profiles suggest that SERPIN1A and IL1RAP may be specifically involved in small airway dysfunction, whereas IgM, CD93 and CCL18 may reflect disease processes less specific for small airway pathology. The post-hoc analysis showed that level of sRAGE, a protein suggested to be a blood based biomarker of smoking induced pathology [32, 33] was found to be lower in PEx from ex-smokers, suggest that PEx samples is capable of reflecting long time effects of environmental challenges, an important feature for sub-phenotyping of disease.

PEx is known to originate to a large extent from the small airway region why we chose to include a group of asthmatic subjects high LCI, an indirect measure of small airway involvement, and severity and level of control in asthma [22]. We found however no differences in protein abundance comparing PEx from asthma patients with or without small airway involvement. This may partly be due to that subjects without small airway involvement were younger than the other groups and that an age independent difference might have been masked by the adjustment for age. Interestingly, we found higher number of proteins to be differentially abundant when comparing the non-asthma group with asthmatics with small airway involvement than with those without small airway involvement, indicating that PEx samples may reflect pathology that drive a more severe type of asthma, also supported by higher ACQ in that group.

The present pilot study was small and primarily dimensioned to highlight the potential of PEx as a non-invasive method for collecting small airway samples compatible with protein biomarker analysis. The quantitative analysis showed however that the described method is sufficiently reproducible and sensitive to allow for detection of shifts in protein abundance profiles. Furthermore, the strong representation of proteins involved in innate immune response, in the PEx proteome, indicates that PEx samples convey pathophysiological information of high clinical value.

Conclusion

Our data illustrate for the first time how non-invasively retrieved RTLF, originating from the small airways in specific, can be analyzed with regard to the relative quantity of over 150 individual proteins. Data reveal that proteins present in PEx to a large extent seem to originate from extracellular vesicles whereof many associated with innate immunity including the complement and coagulation system. Pathological relevance of PEx samples was further demonstrated by showing that the protein found to be differently abundant in subject with asthma with small airway involvement previously are known to be involved in lung disease pathways. Collectively the results indicates that the PExA method provide a novel and non-invasive route to identify novel biomarkers and drug targets contributing to further development of precision medicine in the filed of Respiratory Medicine.

Abbreviations

A-hLCI: Asthma-high lung clearance index

A-nLCI: Asthma-normal lung clearance index

RFU: Relative Fluorescence Units

LOD: limit of detection

GLM: General Linear Model statistics

ANOVA: Analysis of variance

PCA: Principal Component Analysis

GLP: good laboratory practice

PExA: Particles in Exhaled Air, i.e. the method or instrument

PEx: The sample or biological material collected with the PExA method

SAD: Small Airway Dysfunction

ELISA: enzyme-linked immunosorbent assay

LC/MS: Liquid chromatography–mass spectrometry

RTLF: Respiratory Tract Lining Fluid

EF: Enrichment Factor resulting from Gene Ontology enrichment analysis

mHG: minimal hypergeometric model

HG: hypergeometric model

Declarations

Ethics approval and consent to participate

The study was approved by the ethical committee at Gothenburg University, dnr 390-06.

All participants gave their written informed consent to participate.

Consent for publication

All authors have read the last version of the manuscript and approved the manuscript for submission.

Availability of data and materials

The data is owned by AstraZeneca, who financed the extremely expensive analyses. On request, clinical back-ground data and results from performed individual statistical analyses can be retrieved from the corresponding author. Other data are available on reasonable request.

Competing interests

Anna-Carin Olin is reporting competing interests as she is one of the inventors of the PExA method, and boardmember and chairholder of PExA AB. Emilia Viklund is reporting a minor chairhold in PExA AB. Dr Östling reports personal fees from PExA AB during the conduct of the study; and Employed by PExA AB while writing the manuscript but not during the planning and completion of the study.

Funding

Thus study have been funded by AstraZeneca, and supported by The ChAMP (Centre for Allergy Research Highlights Asthma Markers of Phenotype) consortium which is funded by the Swedish Foundation for Strategic Research, the Karolinska Institute, AstraZeneca & Science for Life Laboratory Joint Research

Collaboration, and the Vårdal Foundation. It has also been funded by the Swedish Research Foundation for Welfare (Forte).

None of the funders have been involved in collection of the data. Jörgen Östling was during his employment at AZ involved analysing the data, and Jörgen Östling and Henric Olsson in the interpretation of the data and writing the manuscript, where also Marleen van Geest were involved.

Authors' contributions

Conceived the original idea – AO, MVG, JO

Designed the study – AO, MVG, JO

Planning of the study (groups, methods, ethics, analysis) AO, MVG, JO

Performed data collection – AO, EV

Performed data analysis – JO

Discussed the results - All authors

Supervised the project – AO

Wrote the paper – JÖ and AO with support from the other co-authors

Designed figures and tables – JÖ

Proofreading - All authors

Acknowledgements

We thank Marianne Andersson, Christina Ahlstrand and Helen Friberg for technical support and clinical work.

References

1. Postma DS, Brightling C, Baldi S, Van den Berge M, Fabbri LM, Gagnatelli A, Papi A, Van der Molen T, Rabe KF, Siddiqui S, Singh D, Nicolini G, Kraft M, group As. Exploring the relevance and extent of small airways dysfunction in asthma (ATLANTIS): baseline data from a prospective cohort study. *Lancet Respir Med* 2019: 7(5): 402-416.
2. in 't Veen JC, de Gouw HW, Smits HH, Sont JK, Hiemstra PS, Sterk PJ, Bel EH. Repeatability of cellular and soluble markers of inflammation in induced sputum from patients with asthma. *Eur Respir J* 1996: 9(12): 2441-2447.

3. Farah CS, King GG, Brown NJ, Downie SR, Kermode JA, Hardaker KM, Peters MJ, Berend N, Salome CM. The role of the small airways in the clinical expression of asthma in adults. *J Allergy Clin Immunol* 2012: 129(2): 381-387, 387.e381.
4. van der Wiel E, ten Hacken NH, Postma DS, van den Berge M. Small-airways dysfunction associates with respiratory symptoms and clinical features of asthma: a systematic review. *J Allergy Clin Immunol* 2013: 131(3): 646-657.
5. van den Berge M, ten Hacken NH, van der Wiel E, Postma DS. Treatment of the bronchial tree from beginning to end: targeting small airway inflammation in asthma. *Allergy* 2013: 68(1): 16-26.
6. Telenga ED, van den Berge M, Ten Hacken NH, Riemersma RA, van der Molen T, Postma DS. Small airways in asthma: their independent contribution to the severity of hyperresponsiveness. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* 2013: 41(3): 752-754.
7. Anderson WJ, Zajda E, Lipworth BJ. Are we overlooking persistent small airways dysfunction in community-managed asthma? *Ann Allergy Asthma Immunol* 2012: 109(3): 185-189 e182.
8. Hogg JC, Pare PD, Hackett TL. The Contribution of Small Airway Obstruction to the Pathogenesis of Chronic Obstructive Pulmonary Disease. *Physiol Rev* 2017: 97(2): 529-552.
9. Wright JR. Immunoregulatory functions of surfactant proteins. *Nat Rev Immunol* 2005: 5(1): 58-68.
10. Casals C, Campanero-Rhodes MA, Garcia-Fojeda B, Solis D. The Role of Collectins and Galectins in Lung Innate Immune Defense. *Front Immunol* 2018: 9: 1998.
11. Voelker DR, Numata M. Phospholipid regulation of innate immunity and respiratory viral infection. *J Biol Chem* 2019: 294(12): 4282-4289.
12. Wright SM, Hockey PM, Enhorning G, Strong P, Reid KB, Holgate ST, Djukanovic R, Postle AD. Altered airway surfactant phospholipid composition and reduced lung function in asthma. *J Appl Physiol* 2000: 89(4): 1283-1292.
13. Erpenbeck VJ, Malherbe DC, Sommer S, Schmiedl A, Steinhilber W, Ghio AJ, Krug N, Wright JR, Hohlfeld JM. Surfactant protein D increases phagocytosis and aggregation of pollen-allergen starch granules. *Am J Physiol Lung Cell Mol Physiol* 2005: 288(4): L692-698.
14. Enhorning G. Surfactant in airway disease. *Chest* 2008: 133(4): 975-980.
15. Mead J. The lung's "quiet zone". *N Engl J Med* 1970: 282(23): 1318-1319.
16. Almstrand AC, Bake B, Ljungstrom E, Larsson P, Bredberg A, Mirgorodskaya E, Olin AC. Effect of airway opening on production of exhaled particles. *Journal of applied physiology* 2010: 108(3): 584-588.
17. Larsson P, Larstad M, Bake B, Hammar O, Bredberg A, Almstrand AC, Mirgorodskaya E, Olin AC. Exhaled particles as markers of small airway inflammation in subjects with asthma. *Clin Physiol Funct Imaging* 2017: 37(5): 489-497.
18. Kokelj S, Kim JL, Andersson M, Runstrom Eden G, Bake B, Olin AC. Intra-individual variation of particles in exhaled air and of the contents of Surfactant protein A and albumin. *PLoS One* 2020:

15(1): e0227980.

19. Behndig AF, Mirgorodskaya E, Blomberg A, Olin AC. Surfactant Protein A in particles in exhaled air (PExA), bronchial lavage and bronchial wash - a methodological comparison. *Respiratory research* 2019: 20(1): 214.
20. Bredberg A, Gobom J, Almstrand AC, Larsson P, Blennow K, Olin AC, Mirgorodskaya E. Exhaled endogenous particles contain lung proteins. *Clin Chem* 2012: 58(2): 431-440.
21. Greening NJ, Larsson P, Ljungstrom E, Siddiqui S, Olin AC. Small droplet emission in exhaled breath during different breathing manoeuvres: Implications for clinical lung function testing during COVID-19. *Allergy* 2020.
22. Verbanck S, Paiva M, Schuermans D, Hanon S, Vincken W, Van Muylem A. Relationships between the lung clearance index and conductive and acinar ventilation heterogeneity. *J Appl Physiol (1985)* 2012: 112(5): 782-790.
23. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *The European respiratory journal Supplement* 1993: 16: 5-40.
24. Robinson PD, Latzin P, Verbanck S, Hall GL, Horsley A, Gappa M, Thamrin C, Arets HG, Aurora P, Fuchs SI, King GG, Lum S, Macleod K, Paiva M, Pillow JJ, Ranganathan S, Ratjen F, Singer F, Sonnappa S, Stocks J, Subbarao P, Thompson BR, Gustafsson PM. Consensus statement for inert gas washout measurement using multiple- and single- breath tests. *The European respiratory journal* 2013: 41(3): 507-522.
25. Kjellberg S, Houltz BK, Zetterstrom O, Robinson PD, Gustafsson PM. Clinical characteristics of adult asthma associated with small airway dysfunction. *Respir Med* 2016: 117: 92-102.
26. ATS/ERS recommendations for standardized procedures for the online and offline measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide, 2005. *American journal of respiratory and critical care medicine* 2005: 171(8): 912-930.
27. Juniper EF, Svensson K, Mork AC, Stahl E. Modification of the asthma quality of life questionnaire (standardised) for patients 12 years and older. *Health Qual Life Outcomes* 2005: 3: 58.
28. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 2009: 10: 48.
29. Ding Y, Xu H, Yao J, Xu D, He P, Yi S, Li Q, Liu Y, Wu C, Tian Z. Association between RTEL1 gene polymorphisms and COPD susceptibility in a Chinese Han population. *Int J Chron Obstruct Pulmon Dis* 2017: 12: 931-936.
30. Carnino JM, Lee H, Jin Y. Isolation and characterization of extracellular vesicles from Broncho-alveolar lavage fluid: a review and comparison of different methods. *Respiratory research* 2019: 20(1): 240.
31. Ax E, Jevnikar Z, Cvjetkovic A, Malmhäll C, Olsson H, Rådinger M, Lässer C. T2 and T17 cytokines alter the cargo and function of airway epithelium-derived extracellular vesicles. *Respiratory*

research 2020: 21(1): 155.

32. Sukkar MB, Wood LG, Tooze M, Simpson JL, McDonald VM, Gibson PG, Wark PA. Soluble RAGE is deficient in neutrophilic asthma and COPD. *Eur Respir J* 2012: 39(3): 721-729.
33. Gopal P, Reynaert NL, Scheijen JL, Schalkwijk CG, Franssen FM, Wouters EF, Rutten EP. Association of plasma sRAGE, but not esRAGE with lung function impairment in COPD. *Respiratory research* 2014: 15: 24.

Figures

Fig 1

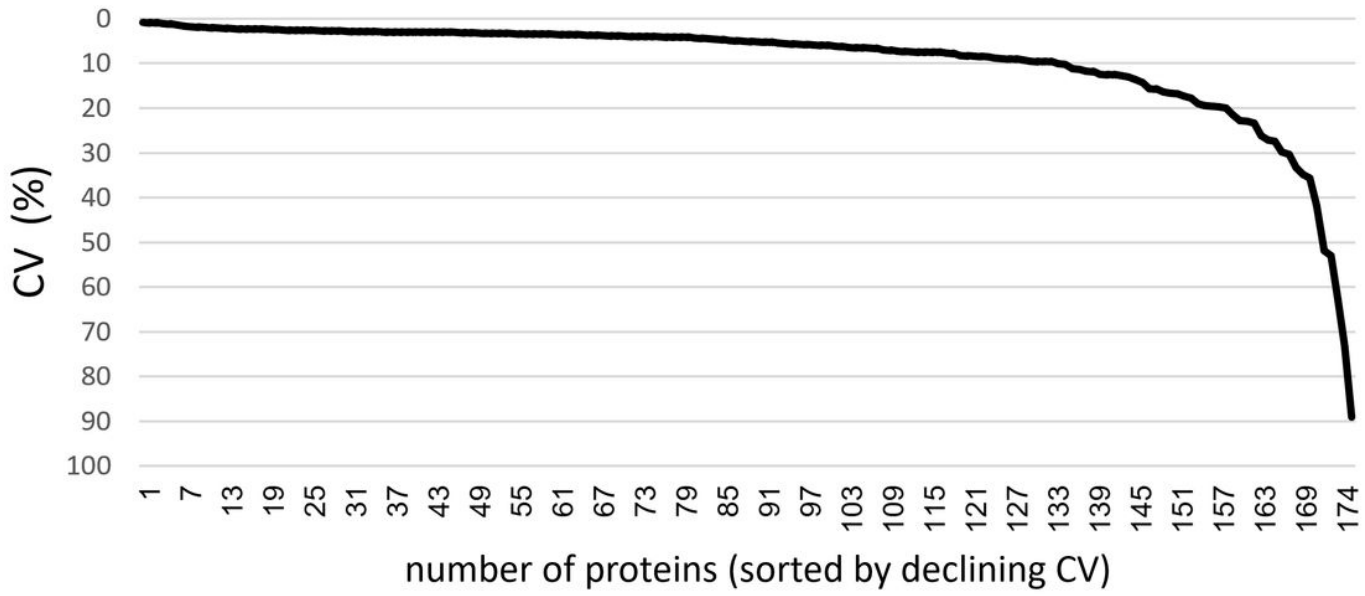

Figure 1

Distribution of CV for 174 proteins detected in a pooled PEx sample (1 µg PEx /ml), analysed 5 times with the SOMAscan 1.1K platform. The pooled sample originated from 6 subjects with asthma and 3 healthy volunteers. Proteins were considered detected if RFU values delivered by SomaLogic were larger than LOD in all 5 replicate samples. Limit of detection (LOD) was calculated as 3 times the standard deviation from the mean RFU signal measured from 3 blank samples.

Fig 2.

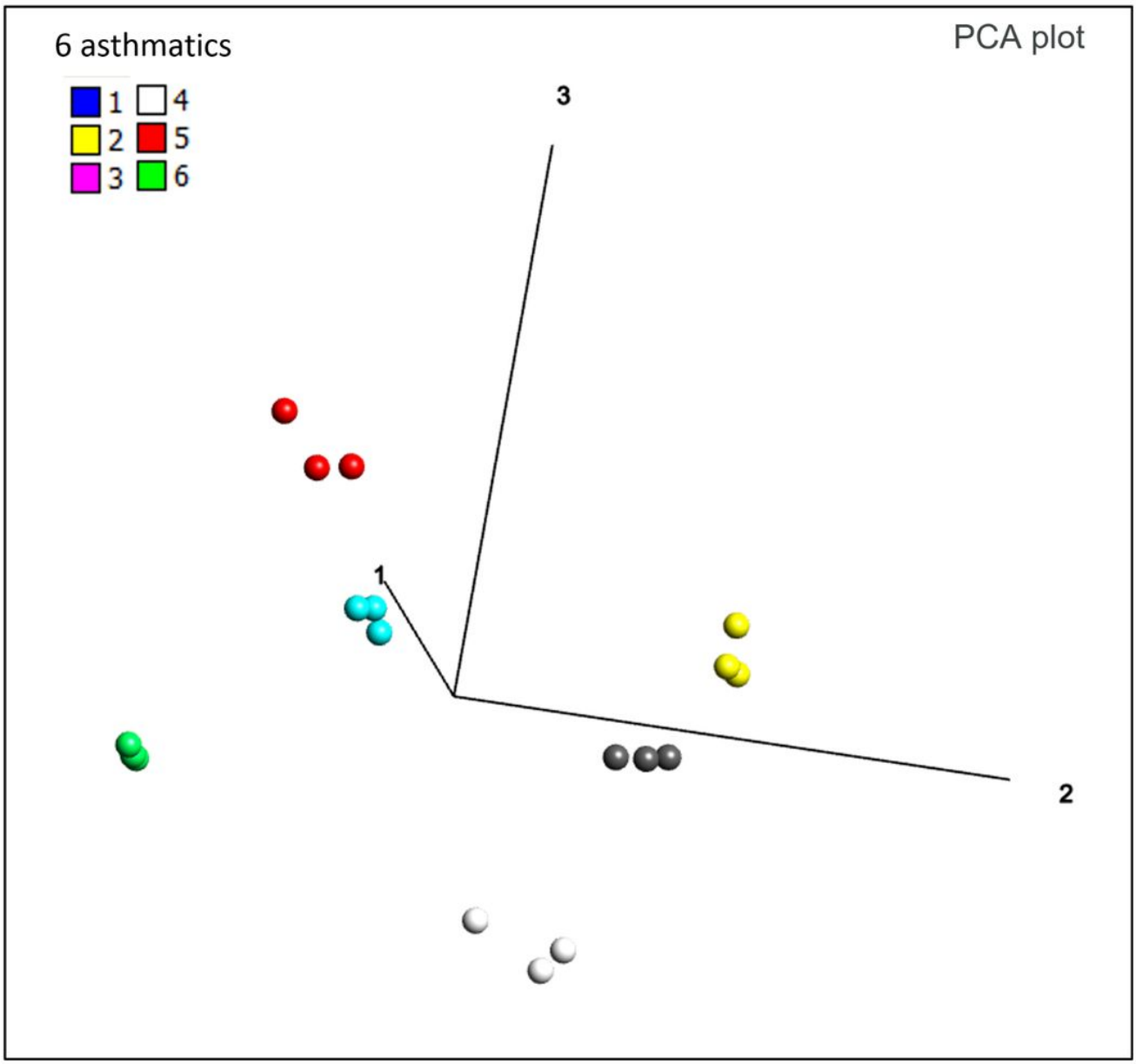

Figure 2

Assessment of intra-individual variability by visual inspection of Principle Component Analysis (PCA) plot. PEx samples from 3 consecutive PEx samples from 6 asthmatic subjects (red, blue, green, white, black and yellow) were analysed with the SOMAscan 1.1K platform. Using ANOVA statistical test based variable selection (q<5.5E-5) 42 out of 114 proteins commonly detected in all 18 samples, were found to discriminate all 6 subjects from each other in a PCA plot, as judged by visual inspection in Omics Explorer software.

Fig 3

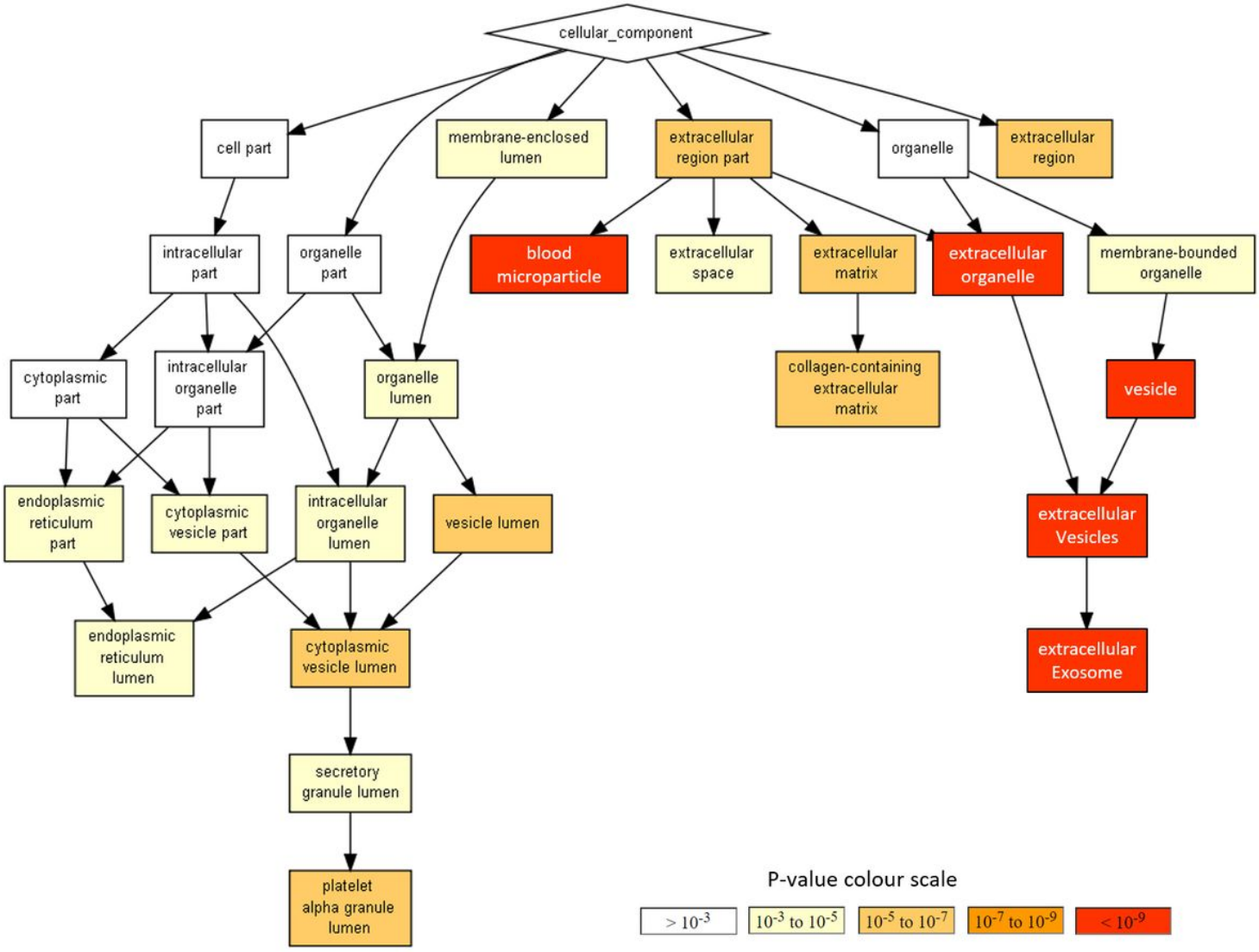

Figure 3

Visualization of results from Gene Ontology (GO) enrichment analysis (GOrilla [28]) matching 207 proteins detected in PEx samples by SOMAscan 1.3K platform, to the GO Cellular Component sub-domain database. Over represented GO terms are organized in a parent-child based hierarchically structure with color-coded significance levels (Fisher's exact test), as indicated in the p-value colour scale.

Fig 4

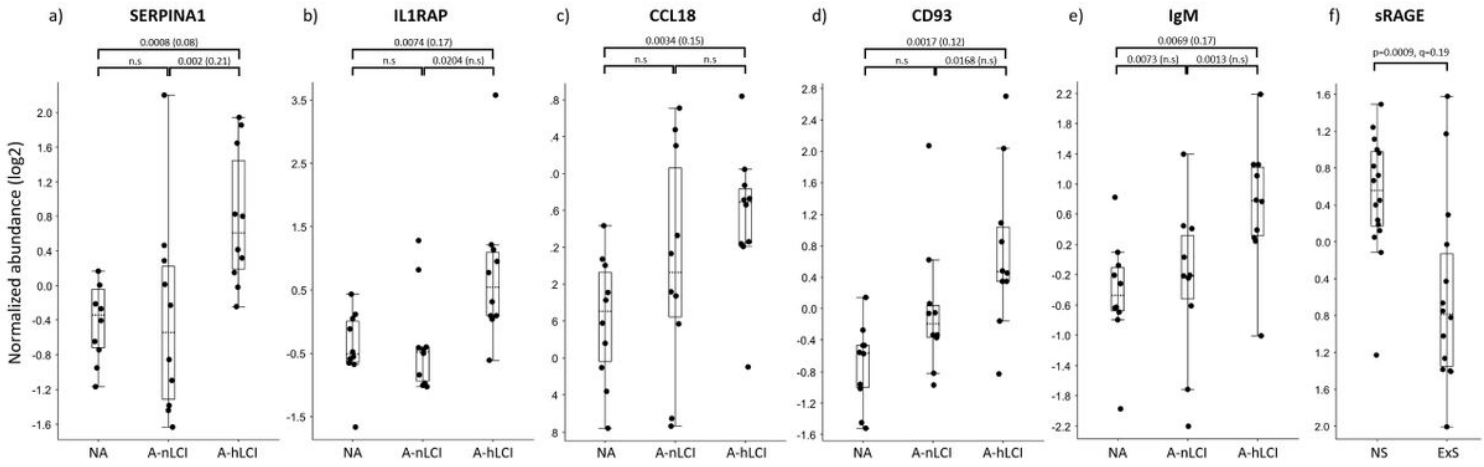

Figure 4

Box plots show examples of SOMAscan data for 6 differentially abundant proteins; a) Alpha-1-antitrypsin (SERPINA1), b) Interleukin-1 Receptor accessory protein (IL1RAP), c) CC motif chemokine 18 (CCL18), d) Complement component C1q receptor (CD93), e) Immunoglobulin M (IgM), in non-asthma (NA), asthma without (A) and with small airway involvement (A-hLCI), and f) Soluble Receptor of Advanced Glycation End products (sRAGE) in never-smokers (NS) and ex-smokers (ExS). Y-axis show normalized abundance (log2 transformation and normalization to mean 0 and variance 1). Box ranges from the 25th to the 75th percentile and median value is marked with dotted line. p values and false discovery rate adjusted p values (q) from various pairwise comparisons are shown over each box plot. Protein abundance data was adjusted for difference in age.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.OstlingetalClinprot211106.docx