**Particles in exhaled Air (PExA): Non-Invasive Phenotyping of Small Airways Disease in Adult Asthma**

**Online Supplement for supporting information**

Marcia Soares, BSc1

Ekaterina Mirgorodskaya, PhD2

Hatice Koca, BSc2

Emilia Viklund, BSc2

Matthew Richardson, PhD1

Per Gustafsson, MD, PhD3

\*Anna-Carin Olin, MD, PhD2

\*Salman Siddiqui, MD, PhD1

1 NIHR Biomedical Research Centre: Respiratory Theme and Department of Infection, Immunity and Inflammation, University of Leicester, United Kingdom.

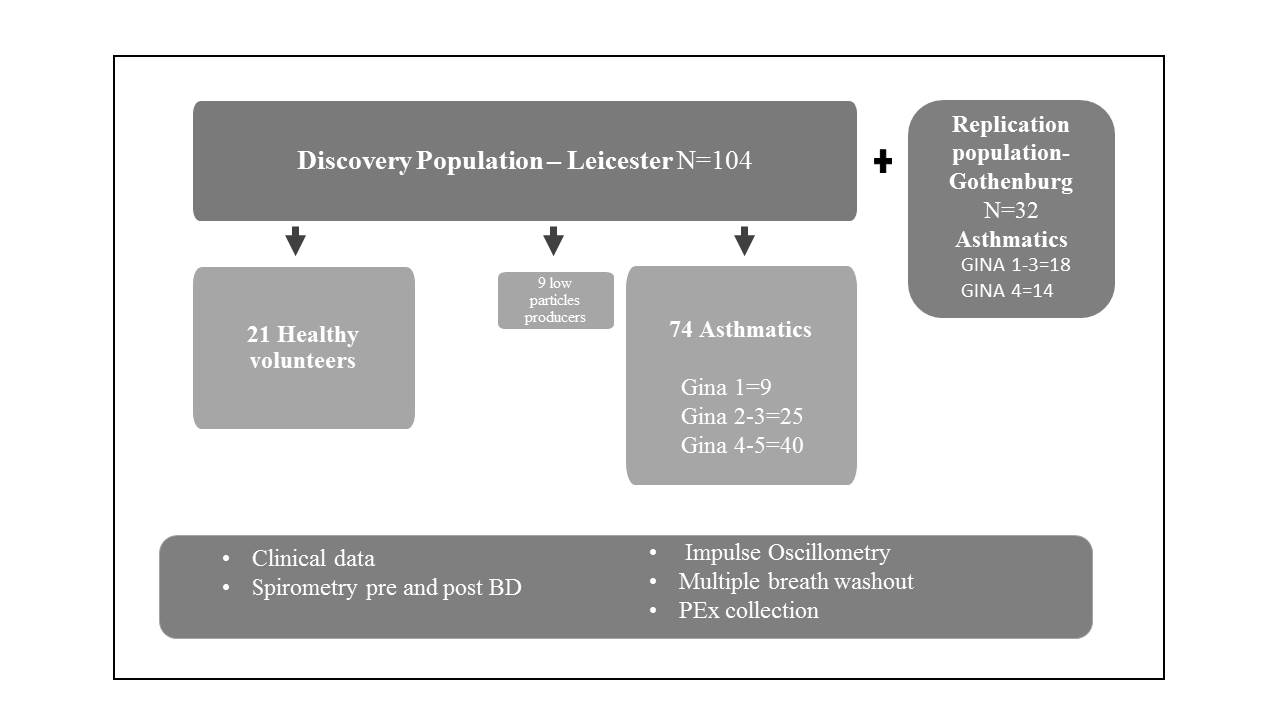
2 Occupational and Environmental Medicine, Department of Public Health and Community Medicine, Institute of medicine, Sahlgrenska Academy, University of Gothenburg.

3 Department of Paediatric Clinical Physiology, Queen Silvia Childrens Hospital, S-416 85 Gothenburg, Sweden

\*These authors contributed equally to the work.

**Data Completeness: Discovery Cohort**

As summary of the discovery and replication cohorts is provided in **Figure E1** in the online supplement. 104 volunteers were screened and recruited from Glenfield Hospital, Leicester UK. 83 asthma patients (Gina step treatment I=9, II-III=25 and IV-V=40) and 21 healthy subjects. Out of 83 asthmatics, 74 were able to produce a PEx sample, 7 had a very low amount of particles per exhalations, therefore, no sample was collected. All controls were able to produce PEx. **Table E1** shows the number of patients from the discovery population undergoing each breathing test.

****

**Figure E1**: Study design and assessments. GINA: Global initiative for asthma.

**Table E1:** Number of patients undergoing each breathing test from the discovery population.

|  |  |  |
| --- | --- | --- |
| Test | Asthma  N=83 | Controls  N=21 |
| PEx collection | 74 | 21 |
| Spirometry | 69 | 16 |
| Impulse Oscillometry | 66 | 17 |
| Multiple breath SF6 washout | 68 | 19 |
| Induced Sputum | 33 | 7 |

**PEx Sampling**

PExA sampling was feasible across the spectrum of asthma severity (**Table E2A and E2B).** After simple explanation and demonstration of the breathing manoeuvre, all patients tolerated repeated airway closure manoeuvres required to generate PEx and there were no adverse reactions e.g. pre-syncope.

Before the sampling session, participants were coached to perform the PExA breathing manoeuvre (**figure E2A and E2B**) which consists of participants first exhaling to residual volume and then holding their breath for 5 seconds. Participants then inhale rapidly to vital capacity and exhale steadily (maximum peak flow of 2000L/s). In between breathing manoeuvres, participants breathed tidally. Subjects performed breathing manoeuvres wearing a nose clip, via a mouthpiece and a two-way, non-rebreathing valve into a box (maintained at 36°C).

Participants were asked to breathe tidally for two minutes prior to PEx sampling in order to prevent contamination from exogenous particles in ambient air. During this time, they inhaled particle-free filtered air via a HEPA filter. The additional two-way valve was closed at this time so all exhaled air during tidal breathing was directed into the room, and no particles from exhaled breath were collected in the impactor.

Subsequently the vacuum pump was turned on commencing the particle sampling. Air was drawn through the impactor at 16L/min from the vacuum pump in order to provide the gas stream through which particles are collected on impactor plates. Clean, particle-free air was added to the reservoir at 18.5L/min to act as a buffer and ensure there was always positive air pressure in the system so no room air enters the instrument.

Participants were instructed to perform the specified breathing manoeuvre repeatedly until a sufficient mass of PEx (50-100 ng) had been collected.

Expiratory flow content was split between a Grimm 1.108 optical particle counter (Grimm Aerosol Technik GmbH & Co, Ainring, Germany) to monitor the number of particles collected in each breath, and an impactor with several stages (modified 3-stage PM 10 Impactor, Dekati Ltd., Tampere, Finland) with a Silicon membrane impaction substrate housed within the terminal impactor (PTFE, diameter 25 mm; Merck Millipore Ltd., Cork, Ireland). The instrument was modified so the exhaled flow could be controlled. At the end of sampling, the sample was transferred to a low-binding Eppendorf polypropylene vial and stored at -80°C until analysis.

**Table e 2a**: PEx sampling fiasibility according to GINA step treatment

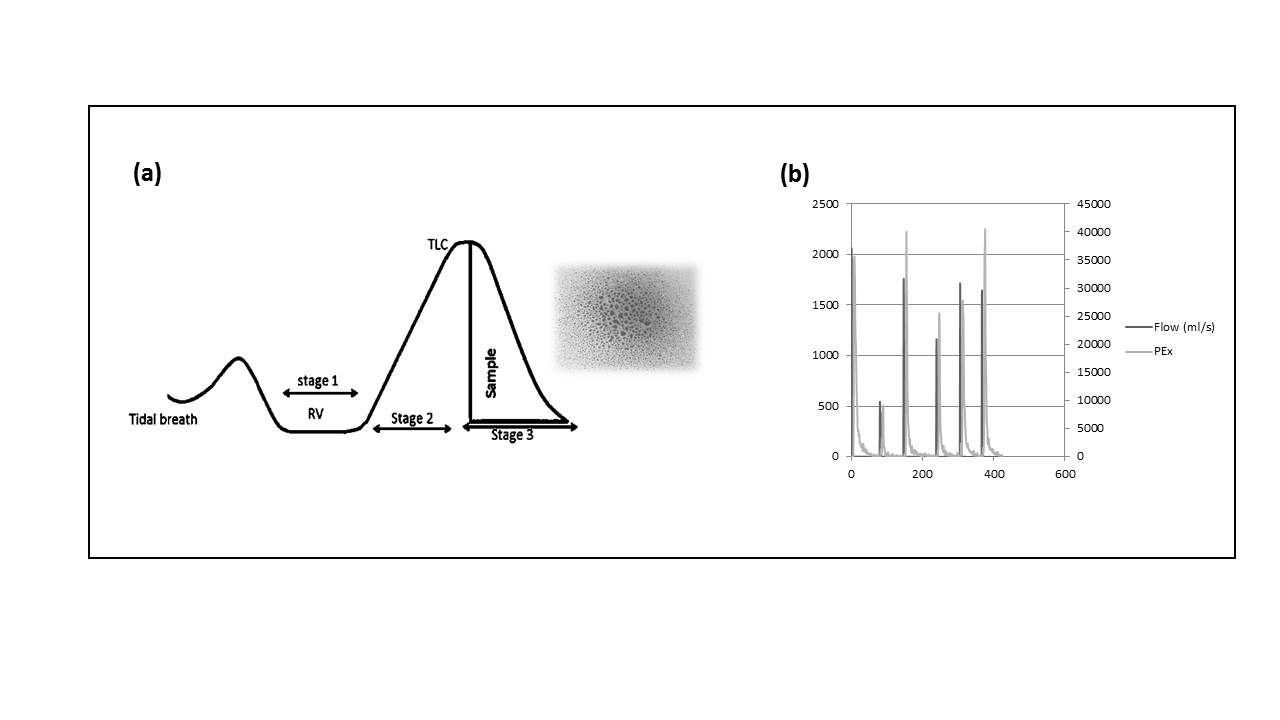
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Healthy | GINA I | GINA II-III | GINA IV-V |
| Patients performing PExA | 21 | 10 | 26 | 47 |
| Sample collected | 21 | 9 | 26 | 40 |
| Sample time (min) | 7 (5) | 10 (6) | 10 (5) | 12 (9) |
| Mean particles per breath | 77294 (61230) | 40609 (26015) | 48281 (30914) | 43144 (42828) |
| Mass PEx acquired (ng) | 104.1 (19.72) | 98.64 (17.35) | 96.73 (22.52) | 94.17 (25.86) |
| Low number of particles (sample not collected)¥ | 0 | 1 | 1 | 7 |
| Mean particles per breath in the low producers‡ | - | 11882 | 10325 | 6322 |

*Definition of abbreviations:* GINA: Global Initiative for Asthma. Data expressed as mean, SD; **¥** Chi-square test: 0.157; **‡**: not enough sample acquired for protein analysis.

**Table E 2B**: PEx sampling fiasibility according to presence or absence of airflow obstruction

|  |  |  |  |
| --- | --- | --- | --- |
|  | Controls | FEV1/FVC ≥ LLN | FEV1/FVC < LLN |
| Patients performing PEx | 21 | 44 | 30 |
| Sample collected | 21 | 44 | 26 |
| Sample time |  | 8 (4) | 15 (9) |
| Mean particles per breath | 77294 (61230) | 53370 (32731) | 40000 (43703) |
| Mass PEx acquired (ng) | 104.1 (19.72) | 98.5 (20.32) | 90.66 (28.25) |
| Low number of particles (sample not collected) | 0 | 0 | 4 |
| Mean particles per breath in the low producers‡ | - | - | 7010 (2857) |

*Definition of abbreviations:* FEV1: Forced expiratory volume in the first second; FVC: forced vital capacity; LLN: lower limit of normal. Data expressed as mean, SD; **‡**: not enough sample acquired for protein analysis.

**Figure E2 a, b**: (a) Breathing maneuver required for PEx sampling. Stage 1: full exhalation till RV and breath hold for 5 seconds; Stage 2: full inspiration till TLC; Stage 3: steady exhalation (flow≈1500ml/s) and collection of PEx material. *Definition of abbreviations:* RV: residual volume; TLC: total lung capacity. (b) Sampling output; sampling flow is represented in black spikes, in milliliters per second (ml/s) and number of particles per breath are shown in grey spikes.

**Sputum results**

Sputum results according to GINA step treatments can be seen on **table E3** and according to the TDA analysis on **table E4**. However, only 42% (40 patients) of the population able to produce a viable PEx sample produced a sample with viability higher than 50%, therefore, sputum results were not imbedded in the main data analysis. Nonetheless, there is a non-significant trend for higher eosinophil counts in the GINA IV-V group, as well as macrophages and sputum epithelial cells, which might indicate damage to the epithelium. Likewise, TDA group 1 (small airway predominant) show higher levels of sputum epithelial cells and macrophages, but no distinction across groups for eosinophils or neutrophils.

**Table E3**: SPUTUM data according to GINA step treatment in the discovery population.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Clinical  Characteristics | Healthy (n=7) | GINA I (n=4) | GINA II-III (n=13) | GINA IV-V (n=16) | Kruskal-Wallis *P* value |
| Total cell count (x10^6/ml) | 4.13 (1.67-7.60) | 4.77 (2.65-25.70) | 7.74 (.346-19.78) | 3.16 (1.39-7.13) | 0.513 |
| Viability (%) | 82 (68-90) | 81 (78-88) | 75 (63-93) | 74 (63-91) | 0.937 |
| Squamous cells (%) | 0.28 (0.00-3.45) | 0.20 (0.00-0.68) | 0.40 (0.00-1.78) | 2.45 (0.00-4.58) | 0.534 |
| Eosinophils (%) | 1.00 (0.00-3.25) | 2.87 (1.81-9.56) | 2.40 (1.50-7.70) | 2.12 (0.56-5.87) | 0.190 |
| Neutrophils (%) | 85.80 (80.50-88.00) | 89.78 (74.20-96.69) | 72.50 (61.55-91.65) | 65.75 (50.33-83.58) | 0.095 |
| Macrophages (%) | 12.00 (9.80-14.00) | 7.75 (1.25-14.63) | 12.50 (4.75-30.95) | 22.50 (11.20-34.13) | 0.147 |
| Lymphocytes (%) | 0.50 (0.00-0.50) | 1.37 (0.25-1.94) | 0.25 (0.00-0.87) | 0.67 (0.31-1.00) | 0.103 |
| Epithelial Cells (%) | 0.50 (0.40-1.75) | 0.25 (0.06-0.62) | 1.50 (0.25-3.12) | 1.87 (0.06-3.12) | 0.103 |

Data expressed as median (Q1-Q3). GINA: Global Initiative for Asthma.

**Table E4** – sputum data according to tda analysis in

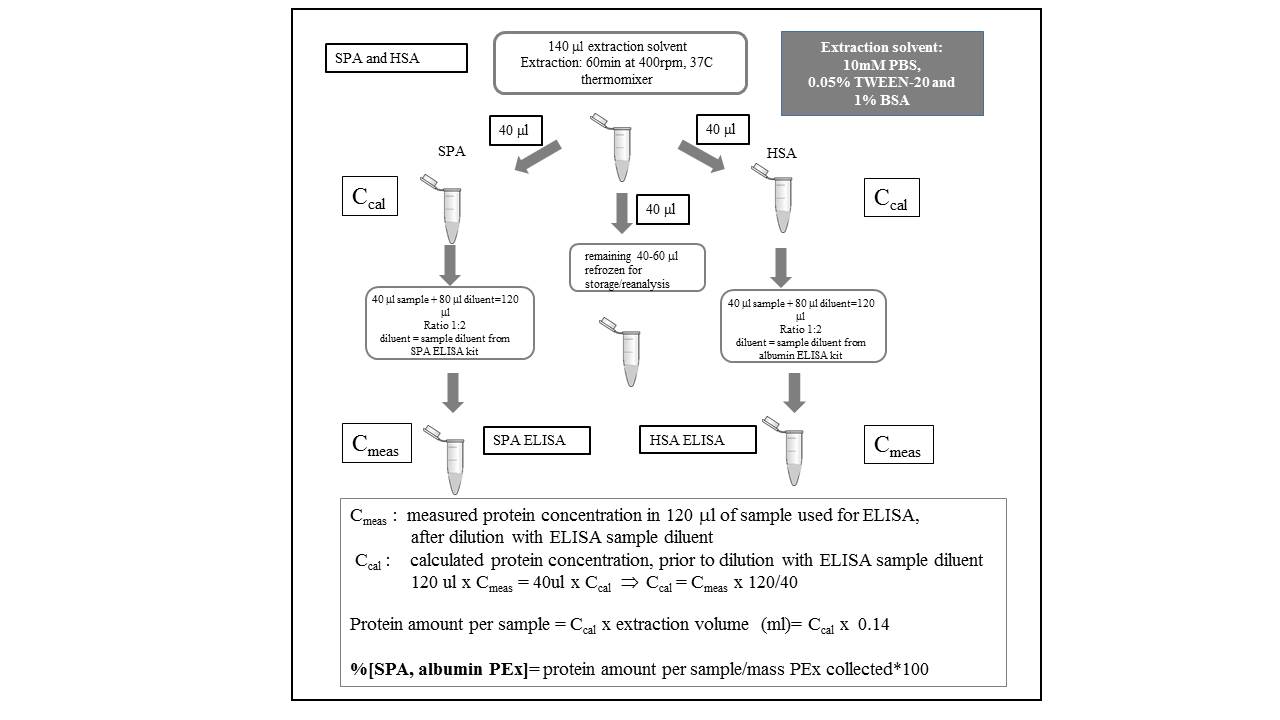
the discovery population.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clinical  Characteristics | TDA group 1  (n=9) | TDA group 2  (n=7) | TDA group 3  (n=22) | *P* value |
| Total cell count (x10^6/ml) Δ | 5.40 (3.59-20.32) | 2.42 (1.03-3.16) | 5.59 (2.11-8.35) | 0.039 |
| Viability (%) | 81 (68-92) | 68 (63-76) | 80 (71-89) | 0.387 |
| Squamous cells (%) | 0.20 (0.00-0.88) | 2.68 (0.00-4.41) | 0.44 (0.30-3.44) | 0.819 |
| Eosinophils (%) | 1.75 (0.50-6.00) | 3.50 (1.25-10.80) | 2.50 (1.12-4.45) | 0.788 |
| Neutrophils (%) | 65.00 (47.50-72.50) | 64.25 (55.85-71.30) | 83.80 (64.10-88.98) | 0.239 |
| Macrophages (%) | 23.00 (10.00-30.00) | 27.50 (17.65-32.50) | 10.50 (7.10-15.88) | 0.097 |
| Lymphocytes (%) | 0.50 (0.00-0.75) | 1.00 (0.60-1.75) | 0.50 (0.25-1.00) | 0.062 |
| Epithelial cells (%) | 2.75 (0.25-3.25) | 2.00 (1.78-4.02) | 0.40 (0.22-1.62) | 0.069 |

Data expressed as median (Q1-Q3). GINA: Global Initiative for Asthma.

**PEx Protein Extraction**

Proteins were extracted from each PEx silicon wafer (**Figure E3**) by adding 140μl PEx extraction solvent, consisting of 0.01M phosphate buffered saline (PBS) (Medicago AB, Uppsala, Sweden), 0.05% TWEEN-20 (Bio-Rad, Hercules, CA, USA), and 1% Bovine Serum Albumin (BSA). The sample vials were spun to ensure that wafers are covered by the extraction solvent and extracted by shaking at 400 rpm for 60 min at 37C in a thermomixer. After locking the wafer between the vial and the lids, the vials were spun again to separate extracted samples from the wafer. The extracted material was pipetted into 3 separate vials of 40μl each: 1 each for SPA and albumin ELISA and the third for storage as a reserve. Each of the 40μl samples for SPA and albumin assays were further diluted with dilution buffer provided in ELISA kits to 120μl. All vials were then stored at -20°C and analysed within a week time period.

**Figure E3**- PEx extraction protocol. *Definition of abbreviations*: SPA: Surfactant protein A; PBS: phosphate buffered saline; BSA: Bovine Serum Albumin.

**PEx protein extraction wortking protocol**

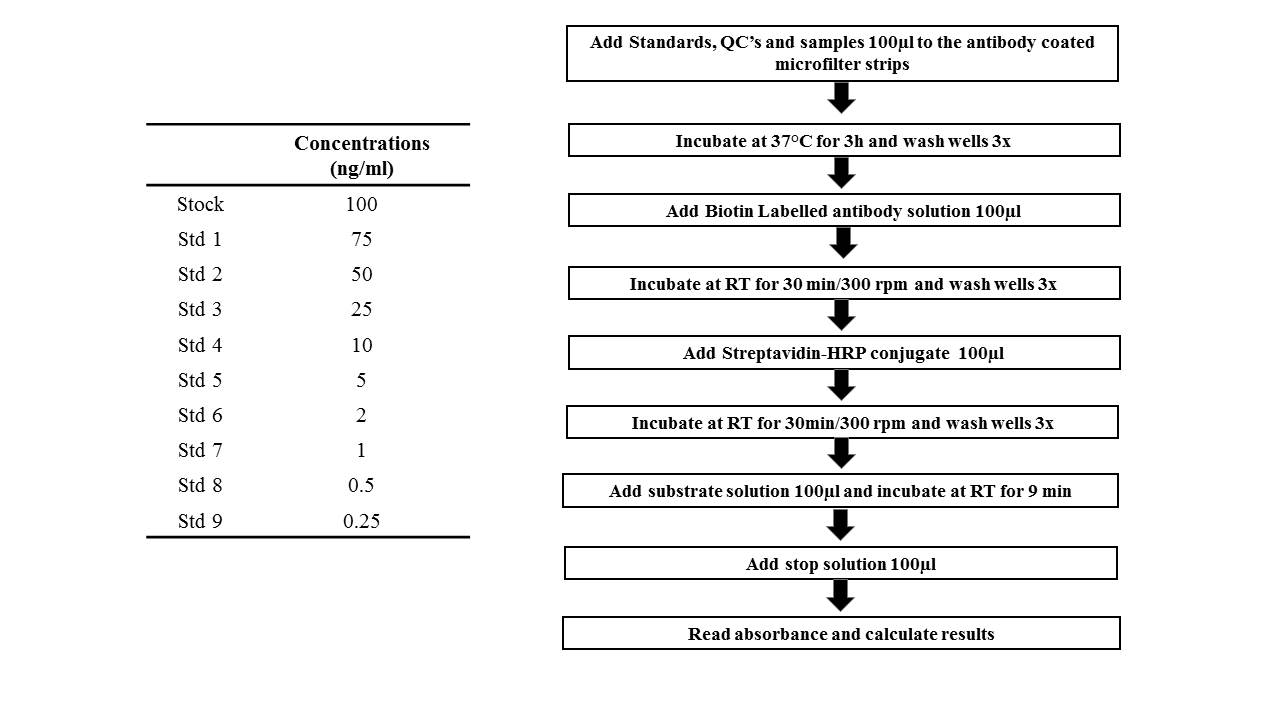
*Required solvents and instrumentation*

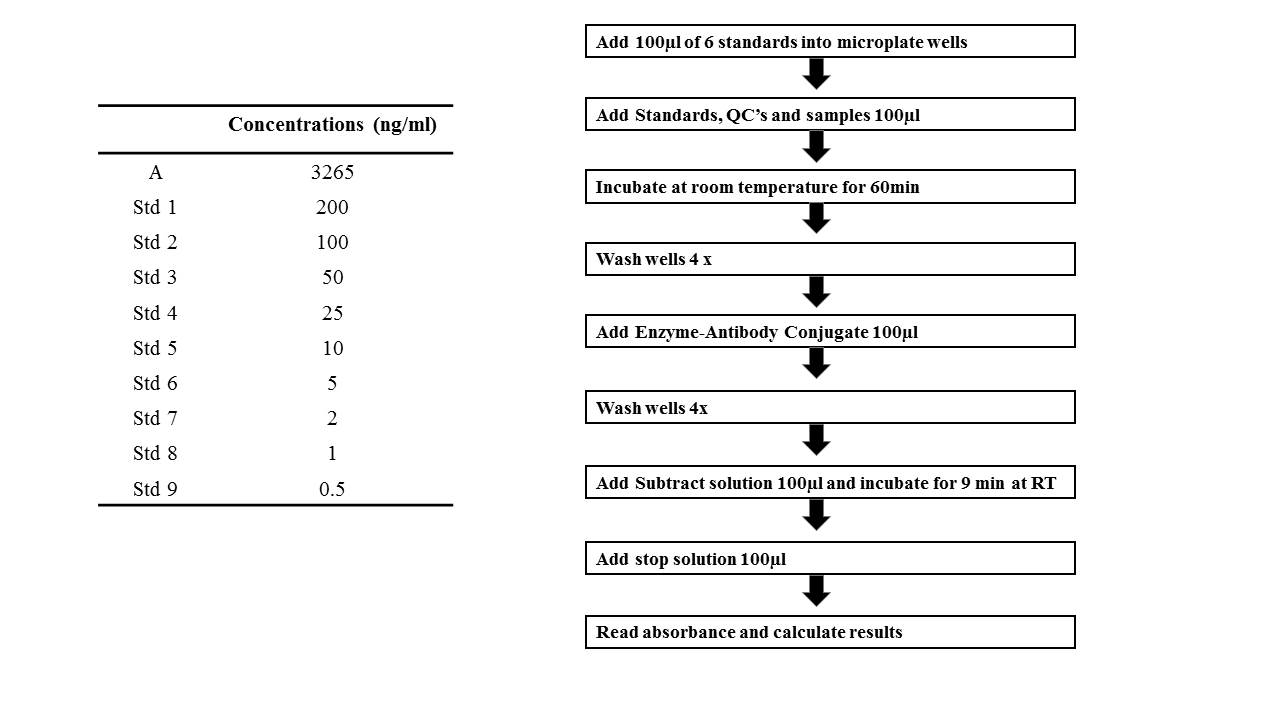
1. PEx extraction solvent: **10mM PBS/ 0.05% TWEEN-20/1% BSA**
2. Sample diluent provided with ELISA kits
3. Thermoshaker

*Procedure*

1. Add 140 μl of PEx extraction solvent to the vial containing the filter, make sure that solvent covers the filter. If needed push the filter gently down with a pipet tip. Do not press the filter too much.
2. Vortex briefly, then spin down briefly
3. Place into thermomixer. Incubate for 60 min, 400 rpm, 37C
4. Spin down vials
5. Pick up the filter edge, bring it up and lock between the vial and the lead. Spin it down for 10 secs to collect all the solvent on the bottom of the vial.
6. Pipette out extracted sample: 3 vials, 40 μl per vial. ( 1 vial: SP-A ELISA, 2 vial: albumin ELISA, 3 vial: reserve)
7. Freeze all three vials
8. Proceed to ELISA according to the corresponding ELISA protocol, with small modifications and extended calibration curve in the low concentration range [as described in **Figures E4 (SPA) and E5 (albumin**)].

Surfactant protein A was quantified using a Human Surfactant Protein A ELISA (BioVendor, Heidelberg, Germany) [E1] and Albumin using a high sensitivity ELISA (ICL, Portland, USA) [E2] according to the product datasheet, with minor modifications.

**Figure E4**- Overview of SPA ELISA, with deviations from manufacterer’s protocol (Adapted). *Definition of abbreviations*: QC: quality controls; Std: standard; RT: room temperature.

**Figure E5**- Overview of albumin ELISA, with deviations from manufacterer’s protocol (Adapted). *Definition of abbreviations*: QC: quality controls; Std: standard; RT: room temperature.

**Evaluation of SPA and albumin: Intra/inter-assay coefficient of variance**

PEx protein analysis in the discovery cohort was performed in 3 different ELISA batches. QC’s sample were prepared to provide values across the quantification range of the assays. The mean concentrations intra-assay and inter-assay coefficient of variation (% CV) for the albumin and SPA quality controls are shown in **table E5**, for high to very low dynamic protein centration ranges in serum and PEx samples in duplicate (from different subjects in each assay).

Serum controls were derived from the same pooled serum sample (mixed unidentifiable individuals) for each of the three assays and demonstrated a between assay coefficient of variance of 0.5% for SPA.

Within assay coefficients of variation for both SPA and albumin in PEx/serum samples were typically less than 10 % across all assays in the discovery cohort.

In the Replication cohort a single SPA and albumin ELISA plate was utilised to analyse all samples and within assay variation was less than 10% for both SPA and albumin in PEx samples (data not shown).

**Table E5:** intra/inter-assay Coefficient of variance for Sp-A and albumin assays in the discovery cohort.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Assay 1 (t=0) | | Assay 2 (t=0+10 months) | | Assay 3(t=0+19 months) | | Inter-assay CV (%) |
|  |  | CV (%) | Mean concentration (mg/ml) | CV (%) | Mean concentration (mg/ml) | CV (%) | Mean concentration (mg/ml) |  |
| Albumin | **QC1** | 2.50 | 78.41 | 1.47 | 81.59 | 1.76 | 69.09 | 8.51 |
|  | **QC2** | 0.23 | 15.10 | 15.26 | 14.60 | 2.99 | 12.41 | 10.19 |
|  | **QC3** | - |  | 8.63 | 2.89 | 7.51 | 2.76 | - |
|  | **Serum** | 7.76 | 75.59 | 2.14 | 76.31 | 2.46 | 76.27 | 0.54 |
|  | **PEx1** | 5.76 | 7.54 | 0.97 | 19.20 | 16.56 | 5.98 | - |
|  | **PEx2** | 0.68 | 5.54 | 7.08 | 8.02 | - | - | - |
| SPA | **QC1** | 1.24 | 6.29 | 0.61 | 6.77 | 6.48 | 3.80 | - |
|  | **QC2** | 4.20 | 1.66 | 11.28 | 1.53 | 3.09 | 1.86 | - |
|  | **QC3** | - | - | 10.09 | 0.39 | 4.89 | 0.66 | - |
|  | **Serum** | 2.88 | 2.39 | 4.25 | 2.60 | 0.76 | 3.09 | 13.40 |
|  | **PEx1** | 2.62 | 7.44 | 8.66 | 5.19 | 11.12 | 3.5 | - |
|  | **PEx2** | 2.83 | 3.29 | 1.82 | 4.05 | - | - | - |

*Definition of abbreviations:* QC1: high quality control; QC2: Low quality control; QC3: very low quality control; CV%: coefficient of variation. Values expressed as CV% between 2 values.

**Replication Cohort**

An additional cohort of 32 asthmatic patient from Skaraborg primary care, Sweden, were added as a replication population, where patients had similar physiological assessments. This cohort was part of a study on small airways dysfunction in adults with asthma [E3], with a total of 196 individuals. We selected 32 individuals based on the R5-R20 values: 1/3 of the individuals with high R5-R20 (around the 90th percentile of discovery population), 1/3 with low values (around the 10th percentile of discovery population), and 1/3 with similar R5-R20 mean values of the discovery.

**Table E6** summarises demographic and clinical characteristics from the replication cohort, divided in asthmatic individuals undergoing GINA step treatment I-III and GINA step treatment IV. The individuals in step IV were significantly older and had a significantly higher absolute value of R5-R20.

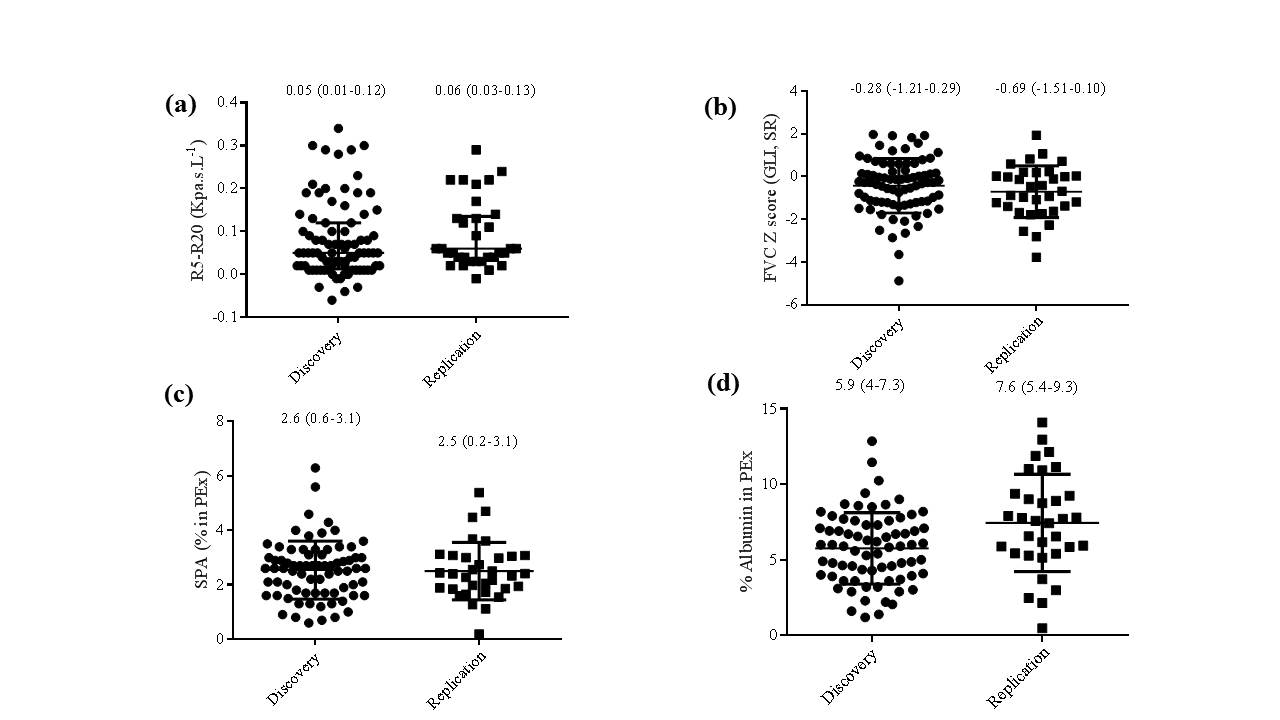
**Table E6:** Demographic and clinical data according to GINA step treatment in the Replication population.

|  |  |  |  |
| --- | --- | --- | --- |
| Clinical  Characteristics | GINA I-III (n=18) | GINA IV (n=14) | Mann-Whitney p value |
| Sex (% male)¥ | 44 | 50 | 0.1 |
| Age (years) | 42 (30-48) | 50 (42-55) | 0.033 |
| BMI (kg/m2) | 24.1 (21.6-28.9) | 27.9 (24-30.3) | 0.342 |
| FEV1/FVC | 73 (68.7-79.9) | 71.9 (62.4-74.8) | 0.419 |
| FVC % predicted | 95.4 (85.9-103.2) | 91.7 (81.7-100.8) | 0.536 |
| FVC Z score | -0.44 (-1.34-0.31) | -0.81 (-1.66-0.06) | 0.536 |
| FEV1 % predicted | 91.6 (77.3-100.5) | 86.2 (66.3-97.3) | 0.338 |
| FEV1 Z score | -0.74 (-1.89-0.04) | -1.33 (-2.88- -0.26) | 0.333 |
| R5-R20 (KPa.s.L-1) | 0.04 (0.03-0.07) | 0.12 (0.06-0.18) | 0.006 |

*Definition of abbreviations:* BMI: body mass index; FEV1= forced expiratory volume in one second; FVC: forced vital capacity.

Data expressed as median, Q1-Q3; **¥** P value based on chi-square test.

**Figure E6 (a), (b), (c) and (d)** shows the median values for R5-R20, FVC Z score, % SPA and % albumin, respectively, in both discovery and replication populations. The difference in the mean levels of % albumin between the discovery and replicate population was statistically significant, as shown in figure **E6 (d)**. However, 52% of the discovery population were patients on GINA step treatments IV-V, against 44% in the replicate data, with no patients on oral steroids.



**Figure E6 a, b, c, d**: **Comparison between discovery and replicate population.** (a) R5-R20, (b) FVC Z score, GLI, SR, (c) % SP-A and (d) %albumin. R5-R20: resistance at 5Hz minus resistance at 20Hz; FVC: forced vital capacity; Data expressed as median and interquartile range.

**Topoligical data analysis**

Topological Data Analysis (TDA) was utilised to evaluate putative small airway phenotypes. Input variables into TDA included physiological measures of small airways dysfunction (Sacin, R5-R20) and PExA protein biomarkers normalised for acquired protein mass, SPA % PEx, and Albumin % PEx.

TDA was performed using the Ayasdi software platform ([ayasdi.com](http://vsea.xuhl-tr.nhs.uk:32224/?), Ayasdi Inc., Menlo Park CA). The central idea in TDA is that the shape of the data has meaning; by understanding the underlying shape of a data set it is possible to discover interesting features such as clusters or subgroups [E4-E7]. TDA generates two dimensional networks of nodes connected by lines and edges to neighbouring nodes based upon patient similarity. Nodes in the network represent clusters of patients and edges connect nodes that contain patients that share phenotypic similarity. Nodes are subsequently coloured by the average value of their respective patients for the variables listed in Figure 3.

Two types of parameters are defined to generate a TDA network:

measurement/notion of similarity, called a metric, which measures the distance between two points in  space, and lenses, which are real valued functions on the data points. Lenses are used to create overlapping bins in the data set, where the bins are preimages under the lens of an interval. Overlapping families of intervals are used to create overlapping bins in the data. Metrics are used with lenses to construct the Ayasdi output. Here, the bins were preimages of rectangles or higher dimensional analogues. Bins were defined by resolution, which determined the number of bins and *gain*, which determines the degree of overlap of the intervals. Once the bins were constructed, we performed a clustering step on each bin, using single linkage clustering with a fixed heuristic for the choice of the scale parameter [E8]. Two nodes were connected if the corresponding clusters contained data points in common. We used two types lenses, the neighbourhood 1 and 2 lens. Theses lenses allowed the embedding of high dimensional data into a two-dimensional space by embedding a K nearest neighbours graph. Similarly to operating a camera, different lens, focus and other setting will generate different shapes and outcomes. The lens function is based on the notion of distance between points in the dataset. In making a TDA graph, the points in the data set are clustered within bins, defined by setting the resolution of the analysis. The metric used was the norm correlation, which is the Pearson correlation coefficient applied to the normalised variables, the resolution and gain where 30 and 3 respectively. To determine how groups of points defined in a TDA graph differ, a non-parametric statistical test (Kruskal-Wallis) or the Chi-Squared test was used to identify parameters that had p<0.05 for either one of the tests, see Tables IIIA, B.

**References**

[E1] BioVendor. Human Surfactant Protein A ELISA Product Data Sheet. Brno; 2009

[E2] Immunology Consultants Laboratory Inc. Human Albumin Immunoperoxidase Assay for Determination of Albumin In Human Samples. Portland, USA, Version3.

[E3] Kjellberg S, Houltz BK, Zetterström O, Robinson PD, Gustafsson PM. Clinical characteristics of adult asthma associated with small airway dysfunction. Respir Med. 2016 Aug 31;117:92-102.

[E4] Siddiqui S, Shikotra A, Richardson M, Doran E, Choy D, Bell A, Austin C, Eastham-Anderson J, Hargadon B, Arron J, Wardlaw A, Brightling C, Heaney L, Bradding P. Airway Pathological Heterogeneity in Asthma: Visualisation of Disease Micro-Clusters using Topological Data Analysis. J Allergy Clin Immunol, 2018 Mar 14.

[E5] Carlsson G. Topology and data. Bull Amer Math Soc. 2009;46(2):255-308.

[E6] Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, Carlsson G. Extracting insights from the shape of complex data using topology. Sci Rep. 2013 Feb 7;3:1236.

[E7] Sarikonda G, Pettus J, Phatak S, Sachithanantham S, Miller JF, Wesley JD, Cadag E, Chae J, Ganesan L, Mallios R, Edelman S. CD8 T-cell reactivity to islet antigens is unique to type 1 while CD4 T-cell reactivity exists in both type 1 and type 2 diabetes. J Autoimm. 2014 May 1;50:77-82.

[E8] Singh G, Mémoli F, Carlsson GE. Topological methods for the analysis of high dimensional data sets and 3d object recognition. InSPBG 2007 Sep (pp. 91-100).