Journal of Breath Research

ACCEPTED MANUSCRIPT

Exhaled breath condensate as a potential biomarker tool for idiopathic pulmonary fibrosis - a pilot study

To cite this article before publication: Barbara Rindlisbacher *et al* 2017 *J. Breath Res.* in press https://doi.org/10.1088/1752-7163/aa840a

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2017 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

source: https://doi.org/10.7892/boris.106594 | downloaded: 19.4.2024

This content was downloaded from IP address 130.92.9.59 on 29/11/2017 at 07:47

Exhaled breath condensate as a potential biomarker tool for idiopathic pulmonary fibrosis- a pilot study

Barbara Rindlisbacher[1*], Carina Strebel[2*], Sabina Guler[3], Attila Kollár[4], Thomas Geiser[3], Georg Martin Fiedler[1], Alexander Benedikt Leichtle[1], Cédric Bovet[1] und Manuela Funke-Chambour[3]

[1]University Institute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, Switzerland

[2]Hospital of Zweisimmen, Zweisimmen, Switzerland

[3]Department of Pulmonary Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

[4]Department of Oncology, Inselspital, Bern University Hospital, Bern, Switzerland

[*]both authors contributed equally to this manuscript

[§]Corresponding author: Manuela Funke-Chambour, MD, Department of Pulmonary Medicine, Inselspital, University Hospital Bern, CH-3010 Bern, E-mail: manuela.funke-chambour@insel.ch. Phone: +41-31-632-2111, Fax: +41-31-632-1404

E-mail: carina.strebel@spitalstsag.ch, barbara.rindlisbacher@insel.ch, sabina.a.guler@gmail.com, Attila.Kollar@insel.ch, martin.fiedler@insel.ch, alexander.leichtle@insel.ch, Thomas.Geiser@insel.ch, cedric.bovet@insel.ch, Manuela.Funke-Chambour@insel.ch

Keywords

Idiopathic pulmonary fibrosis, exhaled breath condensate, mass spectrometry

Accepted Manuscript

Abstract

Idiopathic pulmonary fibrosis (IPF) is a devastating lung disease with poor survival. There is an urgent need to better diagnose and monitor IPF patients as new treatments which slow down disease progression are now available. Exhaled breath condensate (EBC) is easily and non-invasively collected, but analysis of potential biomarkers is difficult due to low concentrations and methodological limitations. We used a non-targeted metabolomics approach to identify discriminatory metabolic profiles that distinguish IPF patients from healthy controls.

For the pilot study set, we collected EBC from 10 stable IPF patients and 10 lung healthy controls. Samples were analyzed by ultra high-performance liquid chromatography coupled to high-resolution mass spectrometry (UHPLC-HRMS) in positive and negative ion mode. After data processing and statistical analysis, 58 metabolites were found to be discriminative between IPF patients and controls in the positive ion mode. One metabolite candidate m/z = 341.3514 at a retention time of 9.94 min was 2.5-fold up-regulated in IPF patients compared to healthy controls and validated in a second set of eight IPF patients and healthy controls. The identity of this metabolic feature still remains elusive.

Our preliminary results identified a distinguished EBC profile of IPF patients compared to controls. Although these results need to be confirmed in additional individuals, EBC sampling for diagnosis and/or monitoring of IPF patients is a promising, new method, which should be further explored. The EBC samples have been obtained within the clinical trial NCT02173145 at baseline.

Accepted Manuscript

1. Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive lung disease affecting mainly men over 60 years [1]. Patients present with unspecific clinical symptoms like worsening dyspnea during exercise and dry cough [1]. Clinical course and prognosis vary, ranging from slow progression to rapid disease deterioration which ultimately leads to death, on average after two to three years [1, 2]. Diagnosis of IPF requires a radiological and/or histological usual interstitial pneumonia (UIP) pattern and exclusion of known causes for interstitial lung disease (ILD), which might impact treatment decisions [1]. Invasive diagnostic procedures, including bronchoscopy and surgical lung biopsy are often needed for a confident diagnosis, but can be associated with high interventional risks, especially in patients suffering from severe respiratory impairment and hypoxia [3]. In addition to diagnostic challenges, the clinical course of the disease is unpredictable [2], treatment options are limited [4-6] and individual treatment response is difficult to assess. Although new medications are in use, which slow down lung functional decline, cure is still impossible and lung transplantation remains the ultimate option for few selected patients [4, 5]. Non-invasive diagnostic and monitoring tools are urgently needed for a safer and more efficient approach to IPF diagnosis, to predict the course of disease and monitor treatment response. A promising solution might be the identification of new biomarkers by exhaled breath analysis. [7].

Exhaled breath condensate (EBC) is a non - invasively collected biofluid of the respiratory tract. EBC contains particles from the airway that enter EBC from the gas phase or from the alveolar lining fluid [8, 9]. Similar to blood or urine, EBC represents a matrix from which biomarkers may be identified. Breath monitoring by EBC sampling is simple and due to its non-invasiveness a particularly promising option to collect biological samples from patients with respiratory impairment. Over the last years, non-invasive methods to diagnose and monitor lung pathologies have gained increasing interest [10].

Accepted Manuscript

In clinical practice NO in exhaled breath of asthma patients is commonly measured to evaluate treatment response [11]. Investigations of EBC in several lung diseases such as asthma [11], chronic obstructive pulmonary disease [12-14], cystic fibrosis [15], lung cancer [16] and interstitial lung diseases [11, 17] are ongoing. Different biomarker patterns in patients with smoking related lung diseases and healthy smokers were identified by analyzing EBC samples [18-21] and an individual metabolomic breath print is postulated [7].

The main advantage of EBC is its non-invasive collection. However, the use of EBC in clinical practice is still limited by the difficulties in analysis of its composition [22]. As EBC contains highly diluted molecules analytic biochemistry assays are difficult to perform. In this study, we investigated EBC by ultra high-performance liquid chromatography coupled to high-resolution mass spectrometry (UHPLC-HRMS), which has been recently suggested as a sensitive and selective technique to identify biomarkers in EBC [23].

2. Materials and Methods

2.1 Study participants

All patients were recruited from the ILD clinic at the Department for Pulmonary Medicine, Inselspital, Bern University Hospital, Bern, Switzerland. Main inclusion criterion for patients was definitive or probable IPF diagnosis [1]. Definitive or possible UIP patterns were diagnosed on high resolution chest CT scans and, if available, UIP was confirmed on lung biopsy. Secondary causes for lung fibrosis were excluded and cases assessed by experienced ILD clinicians, radiologists, and pathologists during multidisciplinary discussion according to the current diagnostic standards [1]. Mortality risk was estimated by the GAP index and staging system, which includes gender, age, forced vital capacity (FVC) and diffusion capacity of the lung for carbon monoxide (DLCO) [24]. Healthy

Accepted Manuscript

volunteers without a history of chronic lung disease were recruited, age ranged between 61 and 86 years. EBC was collected from all patients and healthy controls after they signed informed consent. The EBC samples from IPF patients have been obtained within the clinical trial NCT02173145 at baseline. The study was performed according to human research law and approved by the local ethics committee (KEK 002/14, 246/15).

2.2 EBC Collection

EBC was collected using the Ecoscreen EBC collecting device from Erich Jaeger, Hoechberg, Germany, according to the manufacturer's instructions and the ATS/ERS methodological recommendations for EBC collection [9]. All participants were advised not to eat at least one hour before EBC collection. The cooling device was precooled (-15°C to -20°C). A common blotting paper was placed inside the mouth piece just before the air enters the cooling chamber to reduce contamination with saliva. Participants sat upright, wore a nose clip and performed tidal breathing continuously for 10 minutes. Through cooling and condensation 3-5 ml of breath fluid was produced on average. The samples were placed in aliquots and immediately stored on dry ice at -80 degrees Celsius until analysis.

2.3 Sample preparation

Samples were thawed on ice and 80 µL of -20°C-cold acetonitrile:methanol (ACN:MeOH, 1:1, v/v) containing chlorpropamide as internal standard (1 µg/mL) were added to 20 µL EBC. Samples were shortly vortexed and stored at -80 °C for 30 minutes to precipitate proteins. Afterwards, the samples were centrifuged in two consecutive steps (14'000 *g* at 4 °C for 20 min) to remove precipitates and subsequently, the supernatants were transferred to total-recovery LC-MS glass vials (Waters, Milford,

Accepted Manuscript

MA, USA). A pooled quality control (QC) sample was prepared by mixing 25 µL of individual EBC sample extracts.

2.4 Non-targeted metabolomics analysis

Non-targeted metabolomics analysis was performed on a 2D-UPLC I-Class system coupled to a quadrupole time-of-flight mass spectrometer (Synapt G2-S HDMS, Waters) operated in positive and negative electrospray ionization (ESI) mode. The mobile phases A and B consisted of 0.1% (v/v) formic acid in (A) 1% (v/v) MeOH in H2O and (B) MeOH, respectively. The sample injection volume was 2 µL and the chromatographic separation was performed on an ACQUITY UPLC HSS T3 column (1.0 mm x 100 mm, 1.8 µm, Waters) by applying a solvent gradient from 100% A to 1% A over 11 min starting at 1 min. The solvent flow rate was set to 0.17 ml/min, the column temperature to 50°C, and the autosampler to 6°C. Standard mass spectrometric parameters were 0.5 kV and 50 V for capillary voltage and cone voltage, respectively. Desolvation and source temperature were kept at 450°C and 120°C, respectively. Cone and desolvation flows were set to 150 L/h and 800 L/h, respectively. Leucine-Enkephalin ([M+H]+ m/z = 556.2766, [M-H]- m/z = 554.2620) was acquired every 20 s for lock mass correction. Mass spectra were acquired at a scan time of 0.2 s in the MSe resolution mode over a range of m/z 50-1200. The samples were analyzed with MSe continuum experiments in the positive and negative electrospray ionization (ESI) mode. The pooled QC sample was analyzed with a DDA experiment to obtain MS/MS spectrum information. A system suitability test (SST) containing standards was measured at the beginning and the end of the analytical run to ensure retention time stability, intensity stability and mass error ≤ 8 ppm. The system was equilibrated by injecting 10 times the pooled QC sample and at least 10 QC samples were analyzed in each

Accepted Manuscript

analytical batch. The instrument was controlled via MassLynx (version 4.1, Waters). Representative base peak ion (BPI) chromatograms of a control and a IPF patient EBC sample are presented in figure 1. The pilot study set was analyzed with one replicate per IPF patient and healthy control sample, whereas for the validation set, duplicate analysis was applied.

2.5 Data processing and statistical analysis

In step I, the raw mass spectrometric data were imported and processed with Progenesis QI (version 2.2, Nonlinear Dynamics, Newcastle, UK). After lock-mass correction and chromatographic alignment, ion patterns were deconvoluted between 0.45 – 11 min. As possible ions, $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+H-H_2O]^+$, $[M+H-2H_2O]^+$, $[M+2H]^{2+}$, $[M+3H]^{3+}$, $[2M+H]^+$, $[2M+Na]^+$ and $[2M+K]^+$ were defined in the positive mode, and $[M-H]^-$, $[M-H-H_2O]^-$, $[M-2H]^{2-}$, $[M-3H]^{3-}$, $[M+HCOOH-H]^-$ and $[2M-H]^-$ in the negative mode. Abundances of metabolic features were normalized to all compounds. Noise reduction was performed by excluding ion patterns from further analysis with a 2-fold higher abundance in blank versus QC samples, a peak width < 0.05 min, ion charge > 1 and a coefficient of variation (CV) of the peak area > 30% in the QC samples, as recommended by Dunn and colleagues [25]. Furthermore, the chromatographic window was specified between 0.5 – 10.5 min and low abundant metabolic features (normalized abundance < 100, arbitrary threshold) were excluded. In step II, normalized abundances of the remaining features were subjected to multivariate analysis with SIMCA (version 14, Umetrics, Umeå, Sweden) for biomarker discovery. For the validation set, the features abundances of the duplicate analysis were averaged. After Pareto scaling, principle component analysis (PCA) was performed to visualize trends and detect outliers among observations in the scores plot. PCA was followed by orthogonal partial least square

Accepted Manuscript

discriminant analysis (OPLS-DA) including model cross-validation to compare discriminant features between healthy controls and IPF patients. Results were further integrated if the 10-fold cross-validated correlation R^2(cum) and Q^2(cum) were ≥ 0.5, if the cross validatory ANOVA p-value was ≤ 0.05 and if the permutation test was passed (20 permutations, $R^2 > R^2_{permutation}$ and $Q^2_{permutation} < 0$) [26] [27]. The VIP-plot was used to find discriminative metabolite candidates having a loading vector p(corr) < -0.5 or > 0.5 and a VIP score of > 1.5. All possible ion features were removed if they were not significant (i.e., confidence interval crossed zero) within model cross-validation. In step III, univariate statistical analysis was performed to consistently evaluate the filtered metabolic features. For the pilot study, pairwise comparisons of continuous variables were carried out with arc-sinh transformed normalized abundances in the Progenesis QI software. For the validation study, log10 transformed normalized abundances were used in the SIMCA software. Nominal two-sided p-values < 0.05 were considered statistically significant and corrections for multiple testing were done using the false discovery rate (FDR) method by Benjamini and Hochberg [28].

In step IV, each potential metabolite candidate was reviewed for its ion alignment and chromatographic peak shape. In step V, the selected metabolites were searched against the Human Metabolome Database (HMDB, version 3.6) with a mass accuracy of 8 ppm to obtain potential structure and formula of the metabolite features. Structure elucidation was performed by comparing MS/MS spectra from DDA experiments with data from the open source platforms MetFrag [29] and CSI:FingerID [30] [31].

3. Results

3.1 Study participants

Accepted Manuscript

In total 10 controls and 10 IPF patients were included in the pilot study set. The baseline characteristics of the patients are described in table 1. Mean age of controls and IPF patients was 69 and 68 years with a male predominance (9 out of 10). IPF diagnosis was confirmed histologically in 6 out of 10 patients. Pulmonary function tests showed reduced FVC (mean FVC 67% predicted) and DLCO (mean uncorrected DLCO 43% predicted) in IPF patients. Calculated GAP index (0 to 8 possible score points) ranged from 1 to 6 points (mean 4.2), reflecting GAP stage I to III and a consecutively increased risk of mortality (predicted 3-year mortality 16% for stage I and 76% for stage III) [24].

All except one patient were treated with antifibrotic drugs (Pirfenidone or Nintedanib) for at least three months before study inclusion. All patients and controls were currently non-smokers, most of them with a history of cigarette smoking. Basic physical exam and pulmonary functional tests in controls were normal.

In the validation set, eight IPF patients and healthy controls were included. Healthy control samples were from additional eight individuals. From IPF patients a new sample was taken two weeks after the sample of the pilot study.

3.2 Non-targeted metabolomics analysis

The workflow for measurement, data processing and statistics is illustrated in Figure 2. After data filtration, 1671 and 981 metabolic features were detected by UHPLC-HRMS in the positive and negative ionization mode in the pilot study, respectively. The OPLS-DA models only resulted in acceptable model characteristics in the positive ion mode (table 2). Based on the multivariate analysis,

Accepted Manuscript

26 metabolic features were found to be differentially regulated between healthy controls and IPF patients. With this approach, only metabolic features with strong model contribution and high reliability were selected. Univariate statistical analysis was then performed to consistently screen for additional metabolites. By univariate statistics, 39 metabolic features were significantly regulated in the positive ion mode. By combination of multivariate and univariate approaches, 58 metabolic features were found differentially regulated in the pilot study. The profile of the metabolic features isolated by multivariate and/or univariate statistical analysis revealed a clear discrimination between the healthy and IPF patients (figure 3, list of regulated metabolic features with the corresponding fold changes presented in supplementary file 1).

To validate our results, a new set of samples was profiled by UHPLC-HRMS. Similar to the pilot study, satisfactory multivariate model parameters were only obtained in the positive ionization mode (table 2). 48 discriminative features were selected by multivariate or univariate analysis (list of regulated metabolic features with the corresponding fold changes presented in supplementary file 2). Among them, two metabolic features (*m/z* 410.3731 and *m/z* 341.3514 at retention time 10.06 and 9.94 min, respectively) of the pilot study were confirmed. The feature *m/z* 410.3731 was however regulated in different direction in both study sets, thus only feature *m/z* 341.3514 with a retention time of 9.94 min remained consistently more than two-fold up-regulated in IPF patients compared to healthy controls (figure 4, fold change IPF/healthy of 2.5 in the pilot study and of 2.0 in the validation study, FDR adjusted p-value $\leq$ 0.01). The corresponding extracted ion chromatograms of the validated metabolic feature in the pilot and validation study sets are presented in figure 5.

$C_{21}H_{44}N_2O$ was obtained as potential molecular formula for the regulated metabolic feature. The measured isotope ratio of *m/z* 341.3514 was in excellent agreement with the simulated one of

Accepted Manuscript

$C_{21}H_{44}N_2O$ (98.4% similarity). The major fragment detected by MS/MS experiment (DDA experiment, *m/z* = 128.1076, supplementary figure 1) could be assigned to $C_7H_{14}NO^+$ (with 0.8 ppm error), which could fit with the suggested formula of the parent ion. Based on the accurate *m/z* ratio, possible identities were searched against HMDB and METLIN databases for the validated metabolic feature *m/z* 341.3514. No reasonable metabolite was found against these databases by assuming a protonated [M+H]+ ion. Possible identifications against the databases ChemSpider including PubChem and MassBank are shown in the supplementary file 3. None of these identifications reflect a known biologically relevant molecule.

4. Discussion

In this study, we found 58 metabolic features that discriminated EBC of IPF patients from EBC of healthy controls in a pilot study set. These discriminative features were isolated by UHPLC-HRMS in the positive ionization mode. Among them, one metabolic feature (*m/z* 341.3514 at a retention time of 9.94 min) was up-regulated in IPF patients (FC 2.5) and was validated in an independent set of samples. It is tenable that this metabolic feature is not related to a drug or a drug metabolite, because it was also detected in the healthy controls. $C_{21}H_{44}N_2O$ was found as a potential molecular formula and might represent a potential biomarker for IPF from non-invasively collected biofluids for diagnosis. Further work will be required to identify its structure.

Previous studies on EBC from IPF patients showed promising results. Docosatetraenoyl lysophosphatitic acid (LPA 22:4), isoprostane and H_2O_2 were elevated in IPF compared to healthy controls [32, 33]. In our metabolic study, LPA 22:4 was not detected. This may be due to the lower sensitivity of our methodology compared to the targeted LC-MS/MS assay used by Montesi et al. [32].

Accepted Manuscript

EBC has many advantages over other biofluid sampling methods. Collection is simple, non-invasive, cheap and can be repeated as often as desired. Opposed to induced sputum sampling or BAL it is nearly always safe even in severely respiratory limited patients. These advantages make EBC a promising new tool to investigate lung diseases. Given the possibility of sampling at any point of disease state, EBC can provide a real-time assessment of pulmonary pathologies. Nevertheless, this technique has several pitfalls: collecting and measuring EBC needs to be standardized in order to reduce the effect of confounders on data collection and interpretation and to ensure reproducibility. Recent reviews show high variation in the sampling of EBC due to limited standardization [8, 9]. Some aspects of collecting EBC can be easily standardized for example the material of the collecting tube, the temperature of the collecting device and the duration of collection. However, conditions like ventilation parameters and the degree of dilution vary between subjects and are impossible to standardize but must be taken into account when interpreting and comparing EBC results [10]. We recognize this limitation and further investigations are needed to define the influence of ventilation parameters and the degree of dilution on the molecular profile of EBC. Sampling methods and storage after collection need to be standardized to improve sample reliability [10]. Recently, after submission for this study, a proposed standardization for EBC collection has been published and hopefully will help to improve data about EBC [34]. Overall, poor reproducibility, a lack of large representative studies and the absence of large norm reference cohorts hinder EBC to enter clinical practice [22].

EBC has already been investigated by liquid chromatography mass spectrometry (LC-MS) in several lung diseases, for example in chronic obstructive pulmonary disease (COPD) [13, 21]. Although smaller clinical trials must be confirmed by larger studies, LC-MS seems to be a promising strategy in investigating EBC of other lung diseases [13, 23] [35] and our results contribute to the current body of

Accepted Manuscript

evidence. By UHPLC-HRMS a large amount of data are being generated. Artefact peaks can result in false positives and over-interpretation can bias the results especially in limited sample sets. By validating a single metabolic feature over the 58 regulated features isolated in the pilot study, we demonstrated the importance of a validation set for such an analysis. The discovery of new candidate biomarkers might eventually lead to clinical tools for diagnosis and monitoring of lung diseases. Nevertheless, molecular biomarkers should reflect relevant pathobiological processes and mechanisms of a disease. Using screening methods as we did, carries a risk of detecting markers of physiological processes that have questionable pathobiological importance, and the relevance of many biomarkers in IPF is not yet clearly defined [36]. However, new non-invasive biomarkers are urgently needed.

5. Conclusions

Our study showed that metabolic profiling by UHPLC-HRMS can be applied to investigate EBC as a novel matrix for IPF biomarker discovery. We showed distinctive differences in the molecular profile of EBC from IPF patients compared to healthy controls. Our results should still be confirmed in a larger cohort and further studies might be able to confirm a profile of biomarkers, that could contribute to IPF diagnosis, disease state assessment, prediction of disease progression and response to therapy.

References

1. Raghu, G., et al., *An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management.* Am J Respir Crit Care Med, 2011. **183**(6): p. 788-824.

2. Ley, B., H.R. Collard, and T.E. King, Jr., *Clinical course and prediction of survival in idiopathic pulmonary fibrosis.* Am J Respir Crit Care Med, 2011. **183**(4): p. 431-40.

3. Kolb, M. and Y. Shargall, *Lung surgery in interstitial lung disease-a safe and useful procedure?* J Thorac Dis, 2013. **5**(4): p. 375-7.

Accepted Manuscript

4. Raghu, G., et al., *An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline.* Am J Respir Crit Care Med, 2015. **192**(2): p. e3-19.

5. Funke, M. and T. Geiser, *Idiopathic pulmonary fibrosis: the turning point is now!* Swiss Med Wkly, 2015. **145**: p. w14139.

6. Myllarniemi, M. and R. Kaarteenaho, *Pharmacological treatment of idiopathic pulmonary fibrosis - preclinical and clinical studies of pirfenidone, nintedanib, and N-acetylcysteine.* Eur Clin Respir J, 2015. **2**.

7. Martinez-Lozano Sinues, P., R. Zenobi, and M. Kohler, *Analysis of the exhalome: a diagnostic tool of the future.* Chest, 2013. **144**(3): p. 746-9.

8. Grob, N.M., M. Aytekin, and R.A. Dweik, *Biomarkers in exhaled breath condensate: a review of collection, processing and analysis.* J Breath Res, 2008. **2**(3): p. 037004.

9. Horvath, I., et al., *Exhaled breath condensate: methodological recommendations and unresolved questions.* Eur Respir J, 2005. **26**(3): p. 523-48.

10. Kuban, P. and F. Foret, *Exhaled breath condensate: determination of non-volatile compounds and their potential for clinical diagnosis and monitoring. A review.* Anal Chim Acta, 2013. **805**: p. 1-18.

11. Rihak, V., Zatloukal, P., Chladkova, J., Zimulova, A., Havlinova, Z., Chladek, J., *Nitrite in exhaled breath condensate as a marker of nitrossative stress in the airways of patients with asthma, COPD, and idiopathic pulmonary fibrosis.* J Clin Lab Anal, 2010. **24**(5): p. 317-22.

12. Santini, G., et al., *Exhaled and non-exhaled non-invasive markers for assessment of respiratory inflammation in patients with stable COPD and healthy smokers.* J Breath Res, 2016. **10**(1): p. 017102.

13. Borrill, Z.L., K. Roy, and D. Singh, *Exhaled breath condensate biomarkers in COPD.* Eur Respir J, 2008. **32**(2): p. 472-86.

14. Vestbo, J., et al., *Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE).* Eur Respir J, 2008. **31**(4): p. 869-73.

15. Antus, B., Barta, I., Csiszer, E., Kelemen, K., *Exhaled breath condensate pH in patients with cystic fibrosis.* Inflamm Res, 2012. **61**(10): p. 1141-7.

16. Dent, A.G., T.G. Sutedja, and P.V. Zimmerman, *Exhaled breath analysis for lung cancer.* J Thorac Dis, 2013. **5 Suppl 5**: p. S540-50.

17. Psathakis, K., Mermigkis, D., Papatheodorou, G., Loukides, S., Panagou, P., Polychronopoulos, V., Siafakas, N. M., Bouros, D., *Exhaled markers of oxidative stress in idiopathic pulmonary fibrosis.* Eur J Clin Invest, 2006. **36**(5): p. 362-7.

18. de Laurentiis, G., Paris, D., Melck, D., Montuschi, P., Maniscalco, M., Bianco, A., Sofia, M., Motta, A., *Separating smoking-related diseases using NMR-based metabolomics of exhaled breath condensate.* J Proteome Res, 2013. **12**(3): p. 1502-11.

19. Santini, G., et al., *Electronic Nose and Exhaled Breath NMR-based Metabolomics Applications in Airways Disease.* Curr Top Med Chem, 2016. **16**(14): p. 1610-30.

Accepted Manuscript

20. Malerba, M. and P. Montuschi, *Non-invasive biomarkers of lung inflammation in smoking subjects.* Curr Med Chem, 2012. **19**(2): p. 187-96.

21. Martinez-Lozano Sinues, P., Meier, L., Berchtold, C., Ivanov, M., Sievi, N., Camen, G., Kohler, M., Zenobi, R., *Breath analysis in real time by mass spectrometry in chronic obstructive pulmonary disease.* Respiration, 2014. **87**(4): p. 301-10.

22. Konstantinidi, E.M., Lappas, A. S., Tzortzi, A. S., Behrakis, P. K., *Exhaled Breath Condensate: Technical and Diagnostic Aspects.* ScientificWorldJournal, 2015. **2015**: p. 435160.

23. Beck, O., A.C. Olin, and E. Mirgorodskaya, *Potential of Mass Spectrometry in Developing Clinical Laboratory Biomarkers of Nonvolatiles in Exhaled Breath.* Clin Chem, 2016. **62**(1): p. 84-91.

24. Ley, B., et al., *A multidimensional index and staging system for idiopathic pulmonary fibrosis.* Ann Intern Med, 2012. **156**(10): p. 684-91.

25. Dunn, W.B., et al., *Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry.* Nat Protoc, 2011. **6**(7): p. 1060-83.

26. Triba, M.N., Le Moyec, L., Amathieu, R., Goossens, C., Bouchemal, N., Nahon, P., Rutledge, D. N., Savarin, P., *PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters.* Mol Biosyst, 2015. **11**(1): p. 13-9.

27. Szymanska, E., Saccenti, E., Smilde, A. K., Westerhuis, J. A., *Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies.* Metabolomics, 2012. **8**(Suppl 1): p. 3-16.

28. Benjamini, Y.and Hochberg, Y., *Controlling the False Discovery Rate: A practical and powerful approach to multiple testing.* Journal of the Royal Statistical Society, 1995. **B 57** ((1)): p. 289–300.

29. Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J., Neumann, S., *MetFrag relaunched: incorporating strategies beyond in silico fragmentation.* J Cheminform, 2016. **8**: p. 3.

30. Duhrkop, K., Shen, H., Meusel, M., Rousu, J., Bocker, S., *Searching molecular structure databases with tandem mass spectra using CSI:FingerID.* Proc Natl Acad Sci U S A, 2015. **112**(41): p. 12580-5.

31. Shen, H., Duhrkop, K., Bocker, S., Rousu, J., *Metabolite identification through multiple kernel learning on fragmentation trees.* Bioinformatics, 2014. **30**(12): p. i157-64.

32. Montesi, S.B., Mathai, S. K., Brenner, L. N., Gorshkova, I. A., Berdyshev, E. V., Tager, A. M., Shea, B. S., *Docosatetraenoyl LPA is elevated in exhaled breath condensate in idiopathic pulmonary fibrosis.* BMC Pulm Med, 2014. **14**: p. 5.

33. Chow, S., Thomas, P. S., Malouf, M., Yates, D. H., *Exhaled breath condensate (EBC) biomarkers in pulmonary fibrosis.* J Breath Res, 2012. **6**(1): p. 016004.

34. Horvath, I., et al., *A European Respiratory Society technical standard: exhaled biomarkers in lung disease.* Eur Respir J, 2017. **49**(4).

35. Montuschi, P., Santini, G., Valente, S., Mondino, C., Macagno, F., Cattani, P., Zini, G., Mores, N., *Liquid chromatography-mass spectrometry measurement of leukotrienes in asthma and*

Accepted Manuscript

other respiratory diseases. J Chromatogr B Analyt Technol Biomed Life Sci, 2014. **964**: p. 12-25.

36. Ley, B., K.K. Brown, and H.R. Collard, *Molecular biomarkers in idiopathic pulmonary fibrosis.* Am J Physiol Lung Cell Mol Physiol, 2014. **307**(9): p. L681-91.

37. Xia, J. and D.S. Wishart, *Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis.* Curr Protoc Bioinformatics, 2016. **55**: p. 14 10 1-14 10 91.

Acknowledgments

Liselotte McEvoy, Susanne Krebser, Jeannette Frey (sample collection), volunteers and patients (participation), UHPLC-HRMS analysis was performed at the Clinical Metabolomics Facility, Center of Laboratory Medicine from the Bern University Hospital (Inselspital Bern). Samples from IPF patients were collected within a clinical trial sponsored by the Research fund of the Swiss Lung Association, Berne.

Accepted Manuscript

Figure legends

Figure 1:

Representative base-peak ion (BPI) chromatogram in positive ion mode of an EBC sample from control (upper panel) and IPF patient (lower panel). *Individual differences (Single detection) **Chromatographic region with accumulation of similar as well as differentially regulated metabolites between IPF patients and healthy controls.

Figure 2:

Workflow for sample measurement, data processing and statistical analysis.

Figure 3:

Heat map representing the log10-transformed abundance profile of the regulated metabolic features isolated by multivariate and/or univariate statistical analysis in the pilot study sample set. Identity (summarized by the retention time followed by the corresponding *m/z* ratio or neutral mass n) of the metabolic features are shown on the right side, individual samples in rows, respectively. Cells colored in red represent up-regulated, colored in blue down-regulated abundances. The analysis was done with the MetaboAnalyst online platform [37].

Figure 4:

Distribution of the validated metabolic feature *m/z* 341.3514 eluting at 9.94 min in the pilot (left) and validation (right) study sets (standardized abundance, FDR adjusted p-value < 0.05) between controls and IPF patients in positive ion mode.

Accepted Manuscript

Figure 5:

Extracted ion chromatograms of the metabolic feature *m/z* 341.3514 eluting at 9.94 min detected in the healthy (pink) and IPF (dark blue) patients in the pilot (left) and validation (right) study sets.

Accepted Manuscript

Table 1. Baseline characteristics of controls and IPF patients. Showing either mean ± standard deviation or number and percentage of total.

	Controls	n=10	IPF patients	n=10
Age (years)	68.70 ± 7.91	10	68.44 ± 9.08	10
Men n (%)	9 (90)	10	8 (88.9)	9 (90)
UIP pattern				
HRCT scan	-	-	8	10
Pathology	-	-	6	10
Pulmonary functional tests				
FVC, L	4.14 ± 0.40	10	2.43 ± 0.49	10
FVC (% predicted)	114.60 ± 8.49	10	65.1 ± 11.27	10
DLCO uncorr. (% perdicted)	129.60 ± 31.10	10	43 ±15.02	10
GAP Index score (0-8)	-	-	4.2 ± 1.4	10
Smoking habits				
Current Smokers n (%)	0	10	0	10
Former Smokers n (%)	8 (80)	10	7 (77.8)	9
Pack years	5.50 ± 5.72	8	25.37 ± 15.18	8
Antifibrotic therapy				
Pirfenidon n (%)	-	-	6 (66.7)	10
Nintedanib n (%)	-	-	2 (22.2)	10
No antifibrotic therapy n (%)	-	-	1 (11.1)	10

Accepted Manuscript

Table 2. Characteristics of the OPLS-DA models for the pilot and validation study sets analysed in positive (ESI+) and negative (ESI-) ion mode.

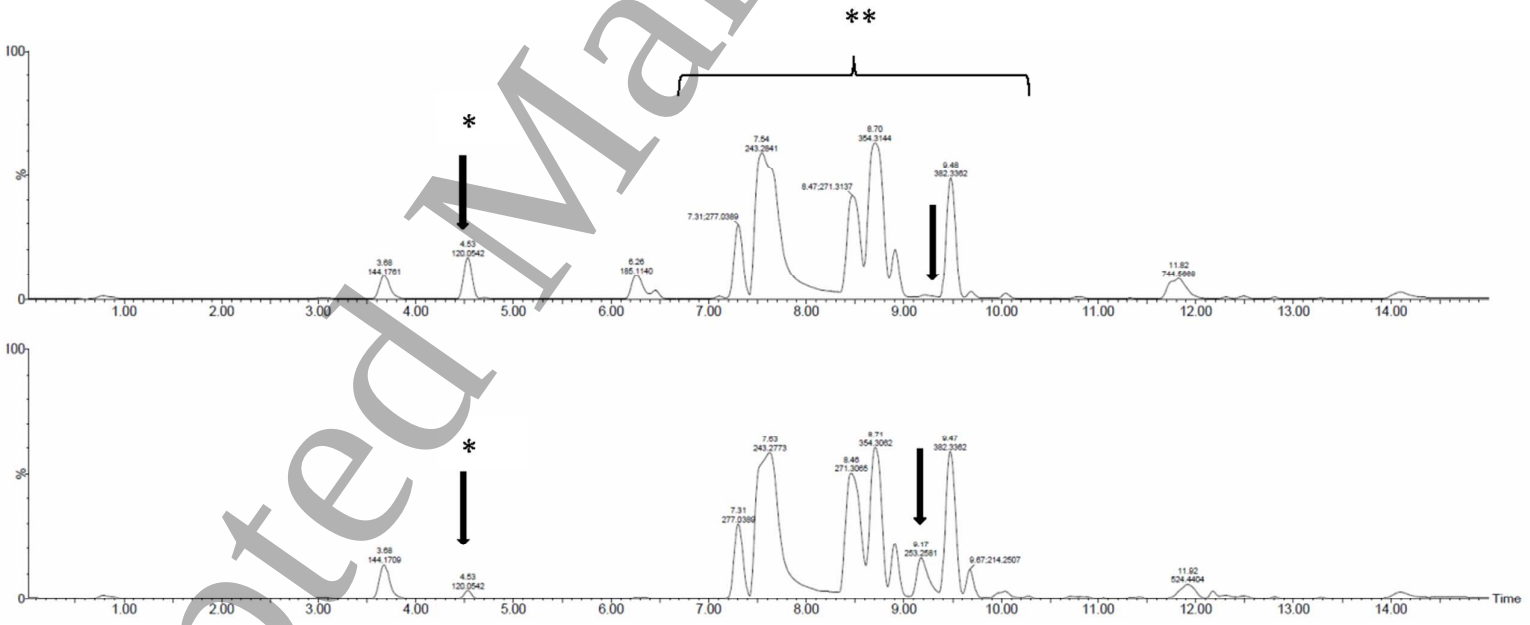
Sample Set	Amount of features	R^2(cum)	Q^2(cum)[a]	Permutation test (n=20)	CV-ANOVA p-value
Pilot ESI+	1671	0.896	0.626	R^2 = 0.535 Q^2 = -0.78	0.03
Pilot ESI-	981	0.605	0.254	R^2 = 0.357 Q^2 = -0577	0.32
Validation ESI+	906	0.693	0.544	R^2 = 0.301 Q^2 = -0.479	0.05
Validation ESI-	220	0.646	-0.112	R^2 = 0.537 Q^2 = -0.41	1

R^2 (cum), cumulative sum of squares explained by all extracted components; Q^2 (cum), cumulative fraction of the predicted variation;

[a] cross validation groups n = 10

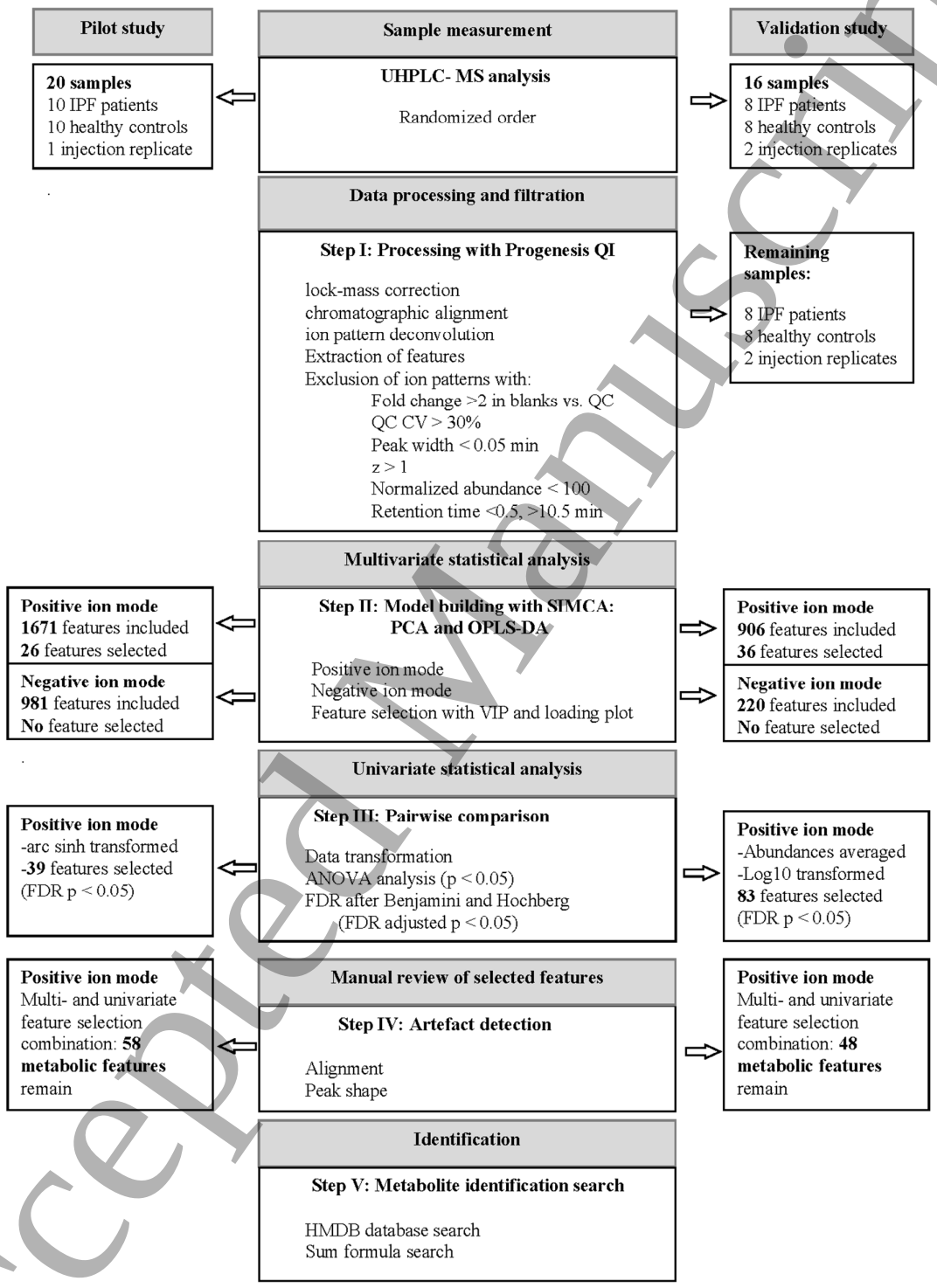
Accepted Manuscript

Figure 1:



Accepted Manuscript

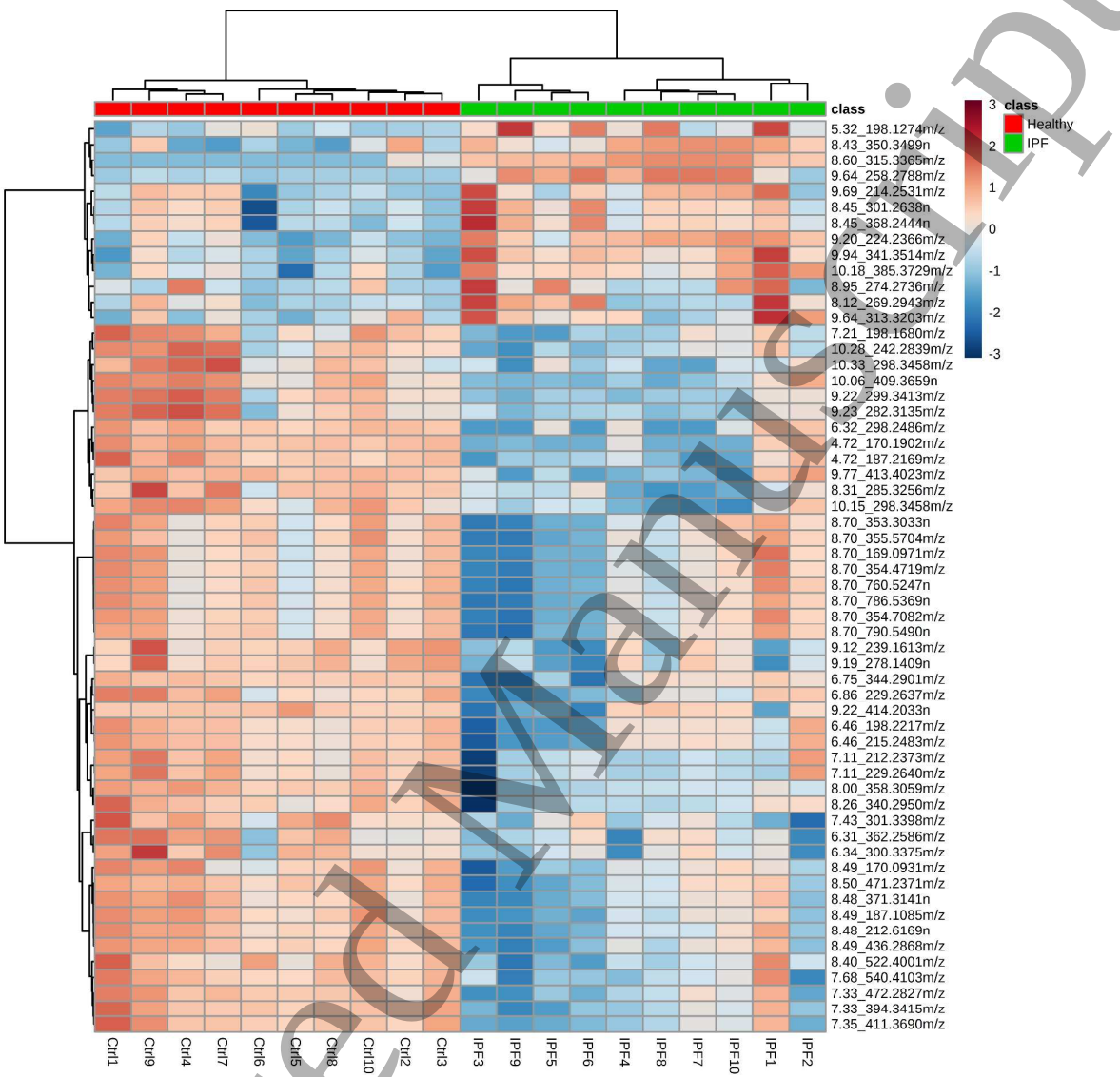
468 **Figure 2:**



469

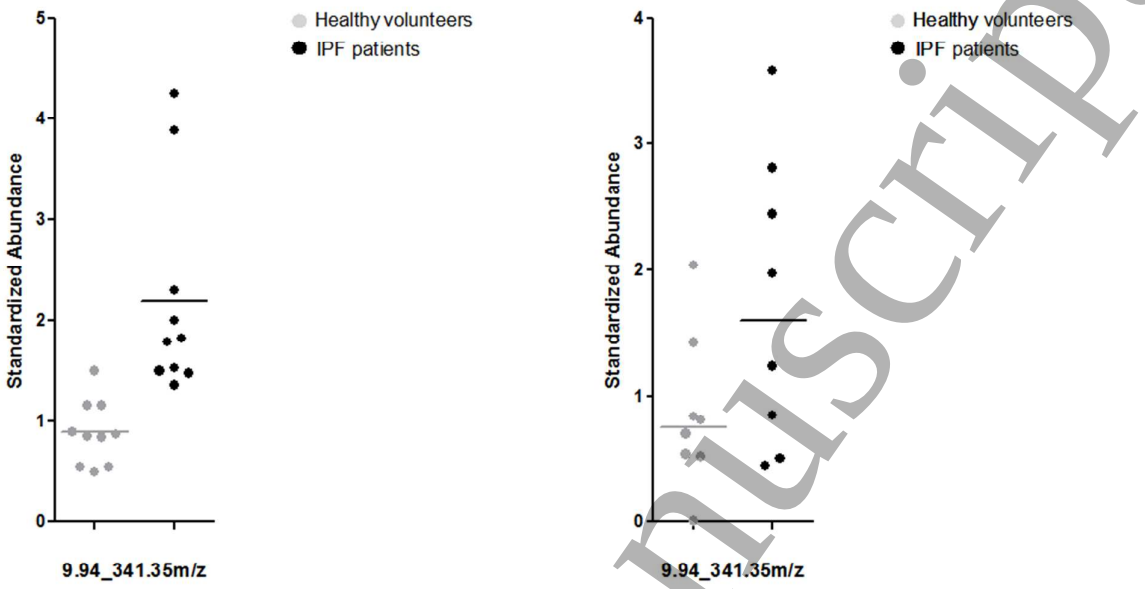
Accepted Manuscript

Figure 3



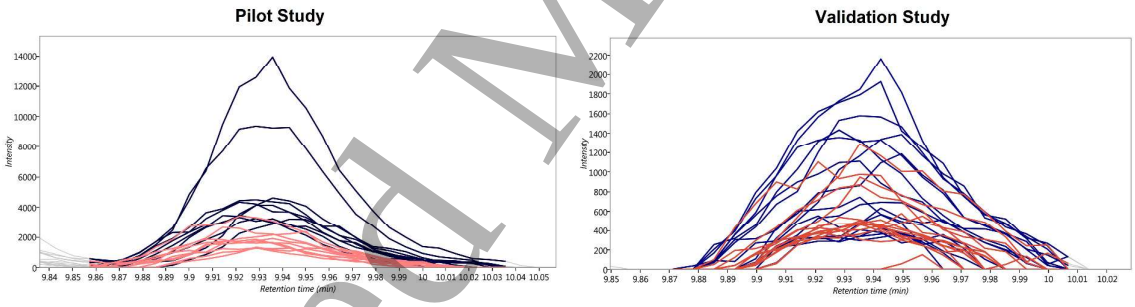
Accepted Manuscript

474 **Figure 4**



475

476 **Figure 5**



477

Accepted Manuscript