A clinical proteomics study of exhaled breath condensate and biomarkers for pulmonary embolism

Gade, Inger Lise; Juul Riddersholm, Signe; Stilling-Vinther, Thomas; Brøndum, Rasmus Froberg; Bennike, Tue Bjerg; Honoré, Bent

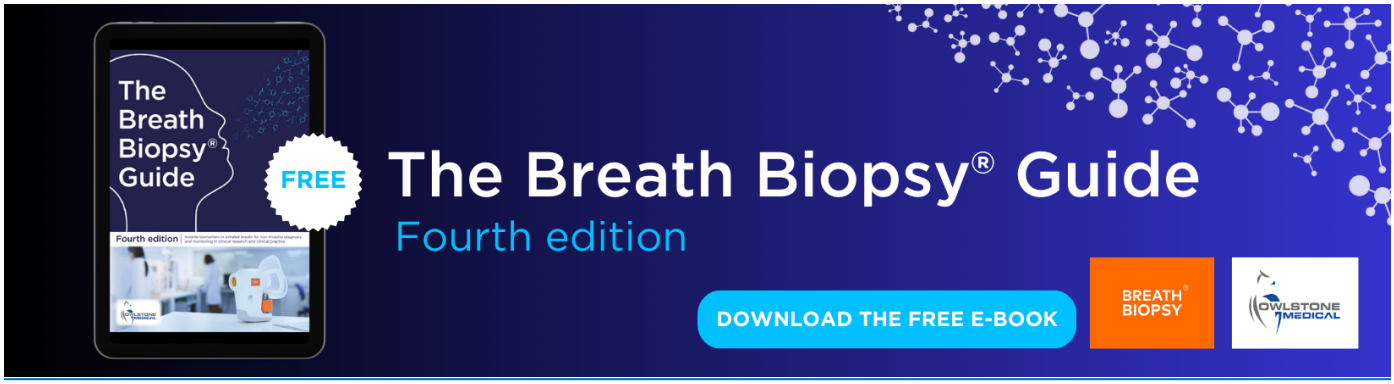
Published in:
Journal of Breath Research

DOI (link to publication from Publisher):
10.1088/1752-7163/ad0aaa

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

A clinical proteomics study of exhaled breath condensate and biomarkers for pulmonary embolism



Journal of Breath Research

PAPER

A clinical proteomics study of exhaled breath condensate and biomarkers for pulmonary embolism

Inger Lise Gade[1,*], **Signe Juul Riddersholm**[2], **Thomas Stilling-Vinther**[3], **Rasmus Froberg Brøndum**[4], **Tue Bjerg Bennike**[5] and **Bent Honoré**[6,7]

[1] Department of Hematology and Clinical Cancer Research Center, Aalborg University Hospital, 9000 Aalborg, Denmark
[2] Department of Cardiology, Aalborg University Hospital, 9000 Aalborg, Denmark
[3] Emergency Care Center, Aalborg University Hospital, 9000 Aalborg, Denmark
[4] Center for Clinical Data Science, Aalborg University and Aalborg University Hospital, 9260 Gistrup, Denmark
[5] Department of Health Science and Technology, Aalborg University, Aalborg, Denmark
[6] Department of Clinical Medicine, Aalborg University, 9000 Aalborg, Denmark
[7] Department of Biomedicine, Aarhus University, 8000 Aarhus, Denmark
[*] Author to whom any correspondence should be addressed.

E-mail: inlg@rn.dk

Keywords: biomarkers, breath tests, proteins, pulmonary embolism, tandem mass spectrometry

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
23 June 2023

REVISED
8 September 2023

ACCEPTED FOR PUBLICATION
7 November 2023

PUBLISHED
17 November 2023

Abstract

Pulmonary embolism (PE) can be a diagnostic challenge. Current diagnostic markers for PE are unspecific and new diagnostic tools are needed. The air we exhale is a possible new source for biomarkers which can be tapped into by analysing the exhaled breath condensate (EBC). We analysed the EBC from patients with PE and controls to investigate if the EBC is a useful source for new diagnostic biomarkers of PE. We collected and analysed EBC samples from patients with suspected PE and controls matched on age and sex. Patients in whom PE was ruled out after diagnostic work-up were included in the control group to increase the sensitivity and generalizability of the identified markers. EBC samples were collected using an RTube™. The protein composition of the EBCs were analysed using data dependent label-free quantitative nano liquid chromatography–tandem mass spectrometry. EBC samples from 28 patients with confirmed PE, and 49 controls were analysed. A total of 928 EBC proteins were identified in the 77 EBC samples. As expected, a low protein concentration was determined which resulted in many proteins with unmeasurable levels in several samples. The levels of HSPA5, PEBP1 and SFTPA2 were higher and levels of POF1B, EPPK1, PSMA4, ALDOA, and CFL1 were lower in PE compared with controls. In conclusion, the human EBC contained a variety of endogenous proteins and may be a source for new diagnostic markers of PE and other diseases.

1. Introduction

Pulmonary embolism (PE) has been dubbed the silent killer because the symptoms vary from asymptomatic presentation of even large central emboli to dizziness, cough, dyspnea, chest pain, hemoptysis, syncope and ultimately acute circulatory collapse/death [1, 2]. A recent study showed that the prevalence of PE in patients with no other symptoms than recent onset of excertional dyspnea is about 20% [3]. The varying and often discrete symptomatology, even shared with other acute cardiothoracic and respiratory diseases can delay the very thought of PE as a diagnosis. Sudden deaths caused by PE are sometimes misinterpreted as myocardial infarction, and PE is among the diagnoses most often missed by clinicians [4, 5]. The standard diagnostic workup and risk stratification includes clinical examination, D-dimer testing, arterial blood gas analysis, electrocardiography, echocardiography and imaging diagnostics such as computed tomography (CT)—and ventilation-perfusion (V/Q) scans. Most of these tests are not specific for PE [6, 7]. The search for biomarkers to enhance prompt diagnosis and improve prognostication of PE is consequently ongoing, mainly focused on markers detected in the blood, but the urine and the volatile organic

compounds of the exhaled breath in humans with suspected PE have also been investigated as potential diagnostic tools for PE [8–10].

Besides volatile organic compounds, the exhaled breath condensate (EBC) contains water vapour with low concentrations of proteins, however sufficient for mass spectrometry analysis. More than 100 proteins have been identified in the EBC, some of which originate from the respiratory tract only, while others are also found in the saliva or upper airways [11–14]. In a recent study, we tested collection of EBC and subsequent proteome analysis in a porcine study of acute PE. Large autologous PE were induced in sedated, anaesthetized, mechanically ventilated pigs [15, 16]. The EBC samples were collected before and after induction of autologous intermediate-high-risk PE. The protein profiles of the EBCs were analysed by label-free quantitative (LFQ) nano liquid chromatography—tandem mass spectrometry (LFQ nLC MS/MS). The proteome of the EBCs collected before PE was compared with the proteome of the EBC collected after PE, and several putative biomarkers of PE were identified [16].

In order to investigate the clinical potential of EBC as a source for biomarkers in humans, this study will compare the protein profiles of EBC from patients with confirmed PE and controls without PE.

2. Methods

2.1. Patients and controls

The power calculation of study sample size was based on the protein that increased most after PE in the porcine study [16], and we aimed at a 5% significance level and a statistical power of 95% (i.e. 5% risk of type 2 errors). The estimated sample size was 26 subjects per group if two groups were compared. Power calculations were performed in STATA IC, ver.14 by the power twomeans syntax.

Patients visited to the emergency department at Aalborg University Hospital between 1 June 2019 and 28 February 2020 with suspected PE were invited to participate upon arrival. Written patient information plus written and oral consent to participate was obtained before EBC collection began. The study was conducted according to the principles embodies in the Helsinki Declaration and approved by the North Denmark Regional Committee on Health Research Ethics (Journal number N-20180086) and registered at ClinicalTrials.gov (Identifier NCT04010760).

A diagnosis of PE was confirmed in case the CT scan and/or V/Q scan confirmed a suspected PE and the patient received treatment for the PE. Patients in whom the diagnosis of PE was rejected were included in the control group. The study was designed to include a healthy control matched on age and gender for each of the confirmed PE cases assuming that we would have at least the protein data quality, that we reached in the porcine study [16]. The matched controls were recruited among out-patients waiting for elective procedures (trans-esophageal echocardiography, radiofrequency ablation for atrial fibrillation or switch of batteries in a pacemaker) at the department of Cardiology, Aalborg University. These subjects had no signs or symptoms of PE at inclusion. In preliminary analyses, we however found that the protein concentration in the clinical EBC samples was below detection limit. Our main focus in this explorative study was to avoid type 2 errors, and we therefore decided to combine the age and gender matched controls and the PE-ruled out patients in one control group in the proteome data analysis.

2.2. Inclusion and exclusion criteria

Patients with clinically suspected PE were invited to participate if they fulfilled the following inclusion criteria: Above 18 years of age possessing legal capacity, being conscious, clinically stable (defined as stable blood pressure and not in need for other treatments) and able to understand written and oral study information. No limits for D-dimer, Wells score or upper age limit were applied. For the age—and sex matched controls we used the following inclusion criteria: Above 18 years of age possessing legal capacity, being conscious, clinically stable (defined as stable blood pressure and not in need for other treatments) and able to understand written and oral study information, same sex and age (within a 10 year range) as an included PE patient and with no signs or symptoms resembling PE. The inclusion started 1st June T2019. The inclusion stopped earlier than the initial planning due to the COVID-19 pandemic.

The following exclusion criteria were applicable for all study participants: a need for intubation/mechanical ventilation or any other organ supporting treatment (e.g. vasopressors, dialysis), being active smokers/vapours or having active cancer (i.e. receiving active anti-cancer treatment or palliative care).

2.3. Collection of EBC

Using the Rtube™ device for collection of EBC, the patients were instructed to breathe through the mouthpiece leading the exhaled breath through a cooling chamber where the vaporous part of the exhaled breath condensed. RTube™ has been used in previous studies of EBC [14, 17–23] including one recently published study where the initial condensing temperature was −80 °C and the collection time was 10 min [24]. We aimed at collecting a total minimum of 3.0 mL EBC per study subject to have enough protein for the subsequent analysis. The porcine studies observed efficient collection for 15 min when pre-cooling the aluminium tube to −80 °C. This gave at least 1.5 ml of EBC per tube[15]. Based on this observation, we aimed at collecting two consecutive EBCs, each of 15 min, from each participant. We pre-cooled the RTube aluminium cooling sleeve to −80 °C and kept it at dry ice during the transport to the patient.

An insulating cover was placed over the aluminium cooling sleeves during the 15 min of EBC collection, but slight temperature increases during sample collection was unavoidable. Since the EBC samples were frozen at the termination of 15 min of collection, the temperature was evidently below 0 °C.

The EBC collections took place whenever suitable between clinical examinations, blood tests and other diagnostic procedures, and did not interfere or delay treatment or diagnostic work-up. The collection of EBC was handled by a trained physician.

2.4. Storage and vacuum centrifugation of EBC samples

The Rtubes™ marked with study subject id containing the EBCs were stored at −80 °C immediately after the collection. The EBC samples were stored until all study subjects were enrolled. The EBC were then set to thaw in batches of four subjects per day. The EBC were transferred to 2 ml Protein LoBind Eppendorf Tubes® and weighed. After determining the volume of EBC by weight assuming that 1 ml equals 1 g, the samples were vacuum centrifugated at 30 °C, 1400 rpm (Eppendorf Concentrator 5301) until dry. The dry product of the EBC samples were kept at −80 °C until sample preparation.

2.5. Sample preparation and quantitative mass spectrometry-based proteomics

Preliminary analysis indicated that the protein concentrations in EBC samples were below 0.5 μg ml^{-1}. Ultimately, the highest sample volume from a single participant was 7 ml with an estimated maximum amount of 3–4 μg of protein which was subjected for LC-MS analysis. The protein concentration was estimated by measuring the fluorescence (excitation at 295 nm, emission at 350 nm) under the assumption that 1 g of protein corresponds to 0.0117 g of tryptophan, as in human and mouse samples[25]. The dry pellets of EBC samples were prepared for bottom-up proteomics using an in-stage Tip digestion kit (PREOMICS, Planegg–Martinsried, Germany). Each sample was prepared according to the manufacturer's instructions. LC-MS-analysis was performed on an Orbitrap Fusion Tribrid mass spectrometer coupled online to a Dionex UltiMateTM 3000 RSLC nano system through an EasySpray™ ion source (Thermo Fisher Scientific Instruments, Waltham, MA, USA). LFQ nLC-MS/MS was performed with the universal method settings essentially as described with few modifications [26]. The peptides were loaded onto a 5 mm C18 trap column (PepMap) and separated over a 50 cm C18 main column. The peptides were eluted from the column at a flow-rate of 300 nl min^{-1} using 99.9% water with 0.1% formic acid, which was mixed with 99.9% acetonitrile with 0.1% formic acid at a gradient increasing the concentration of the latter from 2%–80% over 37 min. The EBC samples were analysed in technical duplicates, with an interval of 4–5 d, except two samples which due to technical issues only could be analysed in singlets.

2.6. Database searches and proteomics data management

The raw data files of the EBC samples were searched against the reviewed UniProt *Homo sapiens* database downloaded 9 February 2020 using MaxQuant (v 1.6.6.0) [27] and used to conduct a relative LFQ analysis of the peptide and protein concentrations. Settings for the database search were as previously described [26] with the match between runs function activated. The LFQ values of each protein were $\log_2$ transformed and technical replicates were averaged by mean. At least one unique peptide was required for successful protein identification. Proteins identified in the reverse database, proteins only identified by site, and potential contaminants, apart from albumin, which based on our porcine studies may be interesting in a PE-context, were removed from the dataset before analysis. The 1% false discovery rate (FDR)-filtered protein and peptide identifications from MaxQuant were further analysed in Perseus (v 1.6.6.0) [28] and *R* (v 4.0.5) [29] using Rstudio v2023.03.0 [30].

2.7. Statistics and bioinformatics

Clinical characteristics of PE patients and controls were presented as mean values and standard deviations or percentages. Means of EBC volumes were compared using the one-way ANOVA, standard deviations were checked using Bartlett's test.

To identify proteins with a PE-associated concentration in EBC, proteins which were not identified in >30% of the EBC samples in the PE or control group after replicate merge were removed. Remaining missing values were ignored and the analyses were conducted using only measured data, except for the principal component analysis (PCA) where missing values were replaced with values from a Gaussian distribution using the impute. MinProb function in the imputeLCMD *R* package [31] ($q = 0.01$, tune.sigma = 0.3) to simulate signals from low abundant proteins. Signal peptide sequence indicative of proteins with active extracellular secretion was identified using SignalP 6.0 [32].

A linear mixed-effects regression model of LFQ-values were fitted a fixed effects of PE-status and a random participant effect, using the lmer function from the lme4 *R* package [33]. *P*-values were calculated with a null-model without PE-status using the ANOVA-function and corrected with Benjamini–Hochberg for multiple hypothesis testing. Proteins were considered to have a significantly different abundance at +/− 0.2 $\log_2$ fold change and p-value < 0.05, as no proteins passed q-value < 0.05. Additionally, a missing-value analysis was performed based on Fisher's exact testing on the data matrix

prior to filtering, as the strict valid value filtering scheme would have removed PE-unique proteins. We used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, stringdb.org) for protein-protein interaction analysis, assessment of biological and molecular function and cellular origin entering gene names for recognition in the data base [34]. Enrichment *p*-values were calculated based on random graph with given degree sequence as described [35].

3. Results

3.1. Study population

In total 157 subjects were screened for study inclusion; 124 patients with suspected PE plus 33 controls with same age and sex as the study participants with confirmed PE. Forty-seven subjects were not included in the study, mostly because of being active smoker ($n = 16$), or because they did not wish to participate after oral study information ($n = 16$), figure 1. Signed consent to participate in the study was obtained for 110 subjects after written and oral study information. PE was suspected in 84 study participants, all included in the study before the diagnostic work-up was completed. Of these, PE was confirmed by conventional diagnostic work-up in 28 cases (PE) and ruled out in 56 cases (nonPE). Controls matched on age and sex were identified for 26 of the 28 PE patients. NonPE patients without pleural effusions or large atelectasis of the lungs were prioritized for proteomics analysis aiming at comparable, homogenous groups. EBC samples from 6 subjects were too sparse for further analysis (EBC volumes 10 μl−1.09 ml) leaving samples from 28 PE patients and 26 + 23 controls for analysis by nano-LC MS/MS. Further details are provided in figure 1.

3.2. Patient characteristics

The PE patients had higher heart rate, higher D-dimer levels, and lower peripheral arterial saturation compared with controls, and more commonly had a body mass index >30. The most common symptoms in patients with suspected PE were dyspnea, chest pain, vertigo, and cough. Interestingly none of the patients reported hemoptysis (table 1). Three patients had a saddle pulmonary embolus, 16 had central PE and 9 were classified as peripheral PEs based on CT or V/Q scans. Nine patients had concurrent DVT and PE, two patients had DVT but no PE. These two patients were included in the control group.

The diagnostic work-up did not result in a specific differential diagnosis in 10 of the 26 nonPE patients, while 7 had either pneumonia, pleuritis or both, 3 were diagnosed with chronic pulmonary diseases or exacerbations of known chronic pulmonary diseases, 2 had reduced ejection fraction, and in 4 patients, the symptoms leading to diagnostic work-up were due to either pneumothorax or muscular tensions.

3.3. Collected EBC volumes

The mean (standard deviation) EBC volume in PE subjects was 4.25 ml (1.22), nonPE 3.99 ml (1.15) and controls 4.11 ml (1.21). The F-test for no difference in the means was 0.31, *p*-value 0.73, which means that there was no difference in EBC volumes (supplemental figure 1).

3.4. Proteome data

A total of 928 proteins identified by at least one unique peptide were detected in the 77 EBC samples after removing proteins identified in the reverse database, proteins only identified by site, and potential contaminants (supplemental table 1). We observed a high variability and many missing values in the proteome data set (depicted in blue, figure 2(a)), with missing values being defined as proteins where a non-zero LFQ value was not obtained, likely due to protein levels below the limit of detection. The average number of proteins identified in each of the EBC samples were 81 (1st quartile: 46–3rd quartile: 97). Eighty-seven proteins were identified in at least 30% of the EBC samples in the PE group or the controls (figure 2 and supplemental table 2) and were submitted for PE-differential analysis. Eleven proteins were identified in the EBC samples from all 77 participants (C3, IGHG1, KRT6, GAPHD, PIP, DSP, ACTG1, DGS1, SERPINB12, CALML5, UBA52).

3.5. Analysis of the exhaled breath protein composition

The three EBC proteins with the highest levels, as indicated by their LFQ values that correlate with absolute concentration, were in descending order: Keratin, type II cytoskeletal 6B (KRT6B), Protein S100-A9 (S100A9), a component of calprotectin, and albumin (ALB) (figure 2(b)). Eigthy-one out of the 87 EBC proteins that were present in at least 30% of the PE or control samples were recognized by the STRING database, and the majority of the proteins had known physical interactions (figure 2(c)). Enrichment analysis revealed that many of the proteins were involved in regulated exocytosis or exosomes. The most significantly represented biological processes and pathways were related to the platelet activation and degranulation, hemostasis, innate immune system, neutrophil degranulation (figure 3). Proteins from many different organs were identified in the EBC, including respiratory system but also the liver, muscles, and intestinal organs (figures 2(c), 3 and supplemental table 3). The presence of a signal-peptide sequence is indicative that the protein is actively secreted from cells, and could be confidently

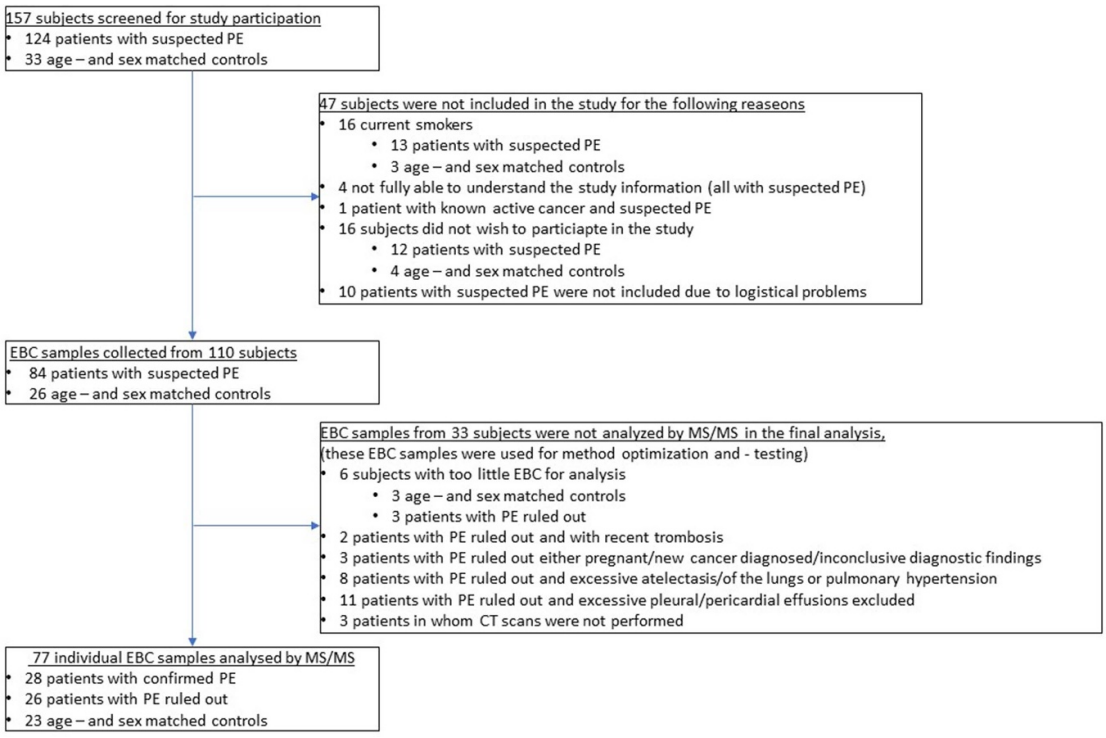


Figure 1. Study recruitment overview.

identified in 29 (33.3%) of the quantifiable EBC proteins. The descriptive analysis indicates that the EBC proteins are of multi-origin, and some maybe transported to the EBC, e.g. from plasma.

3.6. PE-associated EBC protein abundances

The LFQ intensities of the 87 proteins identified in at least 30% of the PE or control samples were investigated with a focus on PE-status. An unsupervised PCA and clustering analysis revealed no obvious clusters of samples based on the available clinical data nor obvious outliers (figures 4(a) and (b), supplemental figure 2), indicative of the main variance in the data likely being explained by a multitude of (unknown) factors. A differential analysis using linear mixed effect models initially identified no proteins passing q-value < 0.05 after multiple hypothesis testing. Given the explorative nature of this study we were first and foremost interested in avoiding type 2 errors, and therefore decided to lower the statistical criteria to include proteins passing uncorrected p-value < 0.05. Seven proteins were found to have significantly different concentrations in PE compared with controls (figure 5, table 2, supplemental figure 4), namely increased levels of HSPA5 and PEBP1 in PE and reduced levels of POF1B, EPPK1, PSMA4, ALDOA, and CFL1. In addition, a unique analysis found that pulmonary surfactant-associated protein A2 (SFTPA2) was significantly more frequently measurable in EBC from PE patients compared with controls (60.7% PE samples vs 34.7% controls, p-value 0.03. supplemental figure 3). To investigate the impact of the control-group, proteins with significantly different LFQ values in PE patients compared with controls were further investigated, by dividing the controls into a 'PE rule out' group and an 'age -and sex matched control' group. Each group was compared to the PE group using linear mixed-effects regression models, correcting for age and sex in the 'PE rule out' group. The directionality of all significantly regulated proteins was retained, regardless if PE was compared to the 'PE rule out', 'age -and sex matched control', or the combined control group (supplemental figure 5). However, only POF1B, CFL1, ALDOA, and EPPK1 retained statistical significance in at least one comparison with the 'PE rule out' or 'age -and sex matched control' groups, which may be attributed to the smaller size of the groups and many missing values, or differences between the EBC compositions of the groups.

4. Discussion

This is the first study to explore the human exhaled breath as a source for diagnostic protein-biomarkers of pulmonary embolism. We identified in total 928 proteins in the EBC using LFQ nLC MS/MS in 77 subjects. The proteins in the exhaled breath represent several distinct biological processes indicating that the air we exhale is a biological sample representing more or less the whole body, similar to a blood sample. The protein concentration in EBC was much lower

Table 1. Clinical characteristics of patients and controls, number (percentage) or when indicated, sd; standard deviations.

Variable, level	Pulmonary embolism ($n = 28$)	Pulmonary embolism ruled out ($n = 26$)	Age -and sex matched controls ($n = 23$)	Total ($n = 77$)	p-value
Male sex	15 (53.6)	10 (38.5)	15 (65.2)	40 (51.9)	0.17
Age, median[iqr]	69 [61.0, 72.8]	52.5 [41.2, 62.0]	67 [57.5, 76.0]	63 [49, 71]	<0.001
Previous VTE	9 (32.1)	7 (28.0)	1 (6.2)	17 (24.6)	0.14
Missing	0	1	7	8	
Lung disease	3 (10.7)	3 (12.0)	2 (25.0)	8 (13.1)	0.56
Missing	0	1	15	16	
Heart disease	2 (7.1)	0 (0.0)	6 (60.0)	8 (12.7)	<0.001
Missing	0	1	13	14	
Anticoagulants	3 (10.7)	3 (12.0)	5 (50.0)	11 (17.5)	0.01
Missing	0	1	13	14	
BMI > 30	12 (42.9)	7 (26.9)	1 (4.3)	20 (26.0)	
Clinical examination					
Heart rate, mean (sd)	90.3 (18.1)	77.7 (17.3)	90.2 (28.7)	84.6 (19.2)	0.03
Missing	0	0	19	19	
Blood pressure:					
Systolic, mean (sd)	143.2 (19.1)	140.3 (19.2)	129.2 (13.1)	140.5 (18.8)	0.25
Missing	0	0	17	17	
Diastolic, mean (sd)	82.6 (13.6)	83.8 (10)	85.7 (11.1)	83.5 (11.7)	0.83
Missing	0	0	17	17	
Laboratory results					
Arterial saturation, mean (sd)	95.1 (4.1)	97.8 (2.3)	100 (NA)	96.5 (3.6)	0.01
Missing	0	0	22	22	
PO2, mean (sd)	8.8 (2.8)	9.1 (4.3)	NaN (NA)	8.9 (3.4)	0.84
Missing	13	15	23	51	
PCO_2mean (sd)	5.5 (2.1)	6.8 (3)	NaN (NA)	6.1 (2.6)	0.08
Missing	2	0	23	25	
Troponin T, mean (sd)	41.6 (60.4)	11.7 (5.2)	NaN (NA)	27 (45.4)	0.02
Missing	5	4	23	32	
D-dimer, mean (sd)	5.5 (5.4)	0.8 (0.7)	NaN (NA)	3.2 (4.5)	<0.001
Missing	1	0	23	24	
Symptoms					
Dyspnea	26 (92.9)	23 (88.5)	0 (0.0)	49 (63.6)	<0.001
Chest pain	13 (46.4)	14 (53.8)	0 (0.0)	27 (35.1)	<0.001
Syncope	1 (3.6)	0 (0.0)	0 (0.0)	1 (1.3)	0.41
Hemoptysis	0 (0.0)	1 (3.8)	0 (0.0)	1 (1.3)	0.37
Vertigo	5 (17.9)	7 (26.9)	0 (0.0)	12 (15.6)	0.03
Caugh	7 (25.0)	5 (19.2)	0 (0.0)	12 (15.6)	0.04
Swollen extremity	5 (17.9)	1 (3.8)	0 (0.0)	6 (7.8)	0.04

than in porcine EBC, and it was, therefore, technically difficult to detect the proteins resulting in a high number of missing values. While the data did not suffice for identifying robust biomarkers for PE, lowering the statistical requirement and including significant proteins passing p-value 0.05 allowed us to demonstrate that biomarkers likely can be identified in EBC as a proof of concept of the idea. We identified eight proteins with significantly different levels in EBC from patients with PE compared with controls. Five of the observed potential diagnostic markers were also identified by discovery-based differential expression analysis in the pre-clinical study of putative diagnostic markers from EBC in a porcine model of acute PE. With p-values < 0.05, the levels of POF1B, EPPK1, PSMA4, ALDOA, and CFL1 were all found to be lower after PE in the porcine study [16].

We identified higher levels of three proteins in PE patients. Surfactant protein A2 (SFTPA2), is a glycoprotein produced by type 2 pneumocytes. Although the mRNA of this protein is found in a few other tissues (prostate, pancreas, trachea), the protein is exclusively expressed in the lungs as opposed to lung surfactant protein D, which is generally expressed on mucosal surfaces [36, 37]. SFTPA2 is actively secreted from the pneumocytes as a component to the pulmonary surfactant to lower the surface tension of the alveoli. Furthermore, SFTPA2 initiates and modulates

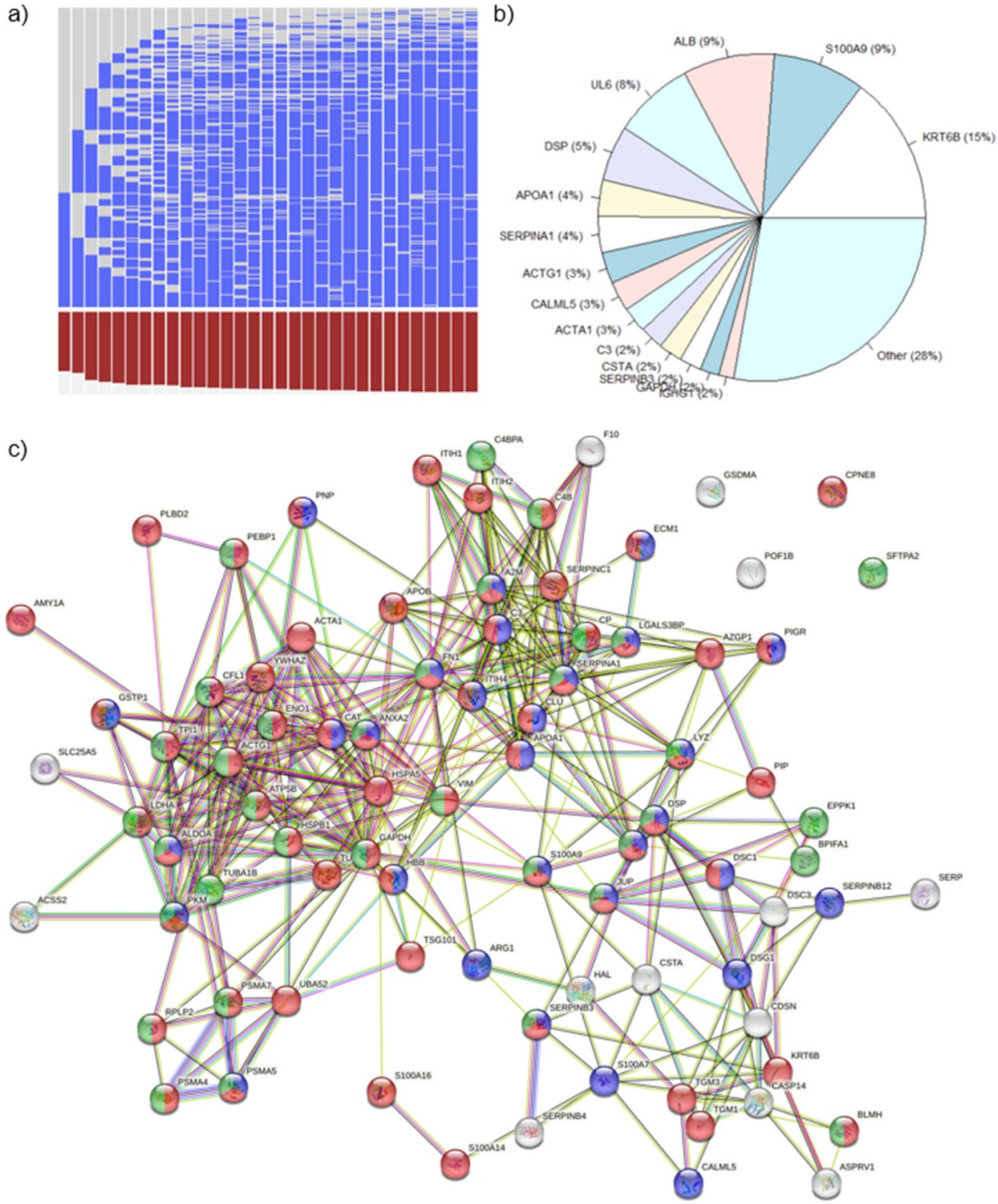


Figure 2. Descriptive analysis of the EBC proteome. (a) Missing value (blue bars) across 31 randomly participants demonstrate a high variability and many missing values across the samples, as expected for the sample type. (b) Major protein components in the EBC. (c) Protein–protein interaction analysis reveals that the proteins detected in the EBC are biologically interconnected. Blue: GO biological process regulated exocytosis (GO:0045055), red. Extracellular space (GO:0070062), green: respiratory system (TISSUES BTO:0000203).

the immune system in the lungs. HSPA5 is highly abundant in type 2 pneumocytes and in immune cells [37]. It is an intracellular chaperone involved in protein transport and storage in the endoplasmic reticulum, and it may inhibit tissue factor procoagulant activity on cell surfaces [38]. PEBP1 is a serine protease inhibitor expressed by most cell types, including type 2 pneumocytes and endothelial cells in the lungs [37]. An array of biological processes are regulated by serine protease inhibitors, including inflammation and coagulation.

The levels of five proteins were lower in PE compared with controls. POF1B is an intracellular protein found in many tissues, the role of this protein is uncertain although it is known to play a role in relation to the cytoskeleton [37]. EPPK1 is expressed in many cell types in the body including alveolar cells [37]. It is an intracellular protein that links to intracellular filaments and controls reorganization to stress. PSMA4 is an intracellular proteasome subunit expressed in most tissues playing a key role in the protein homeostasis [37]. ALDOA is a cytosol protein

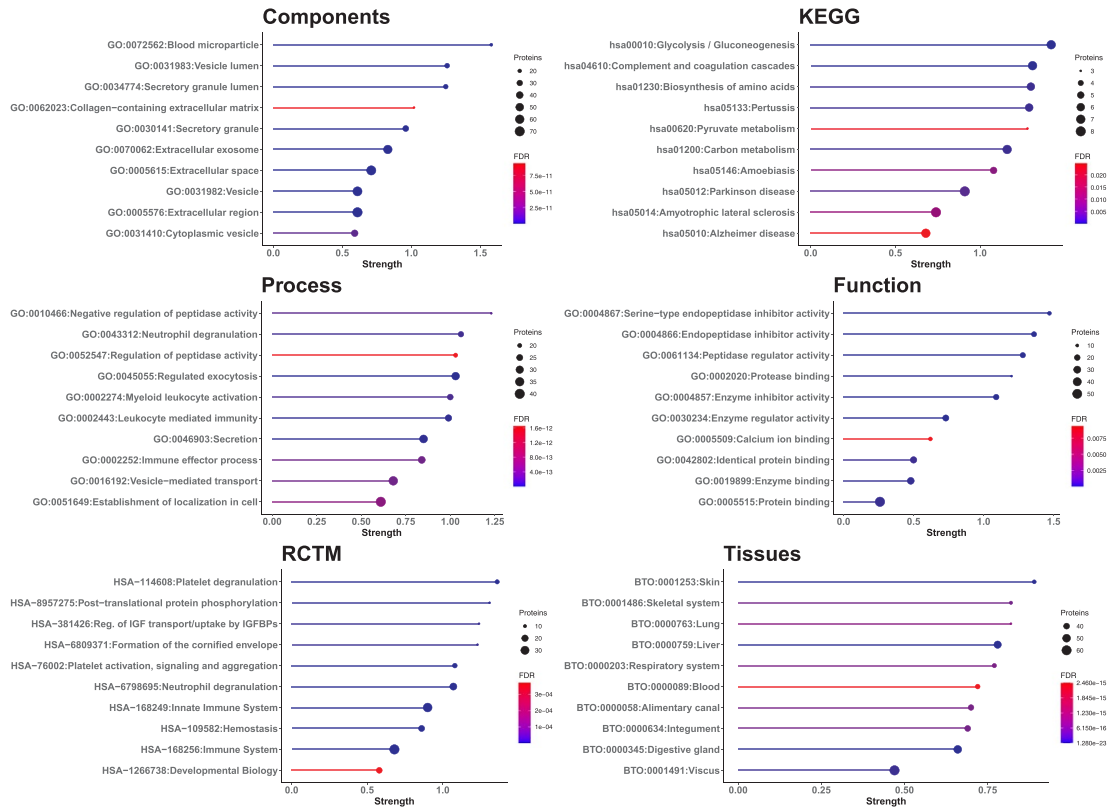


Figure 3. Functional enrichment analysis based on the 87 proteins detected in the exhaled breath condensate. The ordinates depicts the ten most enriched terms with regard to reactome pathways (RCTM), tissue expression, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, the three Gene Ontololy (GO) domains; cellular components biological process and molecular function. The abscissa represents the strength of the enrichment described as Log10 ratio between (I) the number of proteins among the 87 investigated proteins annotated with each term, and (II) the expected number of annotated proteins in case 87 randomly selected proteins were analysed. The colour of each lollipop illustrates the false discovery rate (FDR), i.e. *p*-values adjusted for multiple hypothesis testing by the Benjamin–Hochberg procedure. The dot size illustrates the number of protein involved in each term.

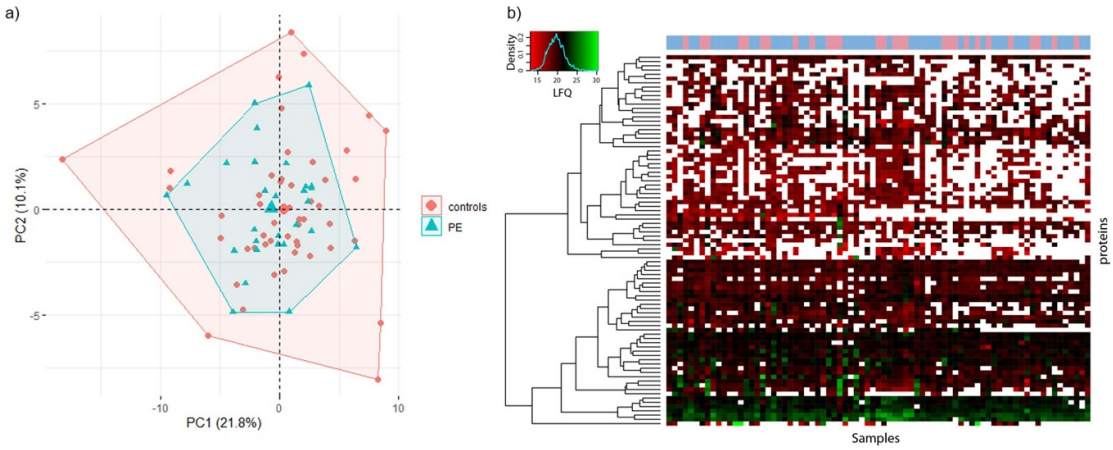


Figure 4. Unsupervised analysis. (a) Principal components analysis (PCA) scores plot of principal component (PC) 1 and 2 visualizing the main variance in the data and separation of pulmonary embolism (PE) and control samples. (b) Hierarchical clustering. LFQ: normalized relative label-free quantitation.

that is important for glycolysis and gluconeogenesis, and it has been reported to protect against damage during hypoxia [39]. CFL1 is a plasma membrane protein expressed in most cell types, particularly in the macrophages. It regulates cell morphology by regulation of the cytoskeleton organization [40].

In summary we see higher levels of proteins that are actively secreted and have effects on the coagulation and inflammation and lower levels of intracellular proteins associated with intracellular 'household' in PE compared with controls.

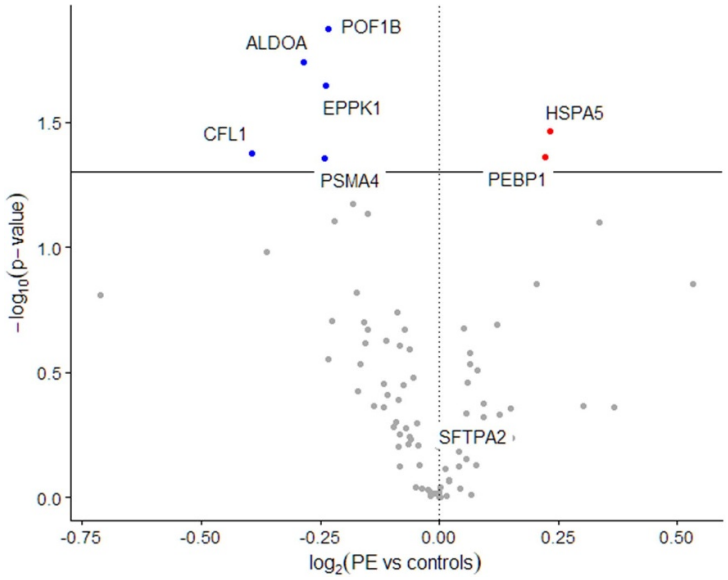


Figure 5. Difference in protein abundances between pulmonary embolism (PE) samples and controls. *X*-axis represents the averaged abundance difference of each protein (circles) between PE- and control samples. The statistical significance is given on the *y* axis, and the black line represents the significance cutoff, with p-value < 0.05. Blue proteins are present at lower levels in the PE group, and red proteins are higher.

Table 2. Proteins with significantly different LFQ values in PE patients compared with controls based on mixed linear effect models.

Protein names	Gene name	Unique protein ID	Fold-change (log2)	*p* value
78 kDa glucose-regulated protein	HSPA5	P11021	0.23	0.035
Phosphatidylethanolamine-binding protein 1	PEBP1	P30086	0.22	0.044
Protein POF1B	POF1B	Q8WVV4	−0.23	0.014
Epiplakin	EPPK1	P58107	−0.24	0.023
Proteasome subunit alpha type-4	PSMA4	P25789	−0.24	0.044
Fructose-bisphosphate aldolase A	ALDOA	P04075	−0.28	0.018
Cofilin-1	CFL1	P23528	−0.39	0.042

Although we performed a power calculation based on the porcine study [16], we found that the lower protein content in the human EBC samples was several-fold lower than in porcine EBC samples which was a large challenge and most proteins were not detected in all samples, resulting in a lower statistical power to detect protein changes. As described, the nonPE patients were relatively healthy, and the EBC volumes and protein profiles of the age—and sex matched controls and the nonPE patients were comparable (supplemental figures 1and 2). We therefore combined the nonPE and C groups to increase the sample size of the control group and the power of the study, eventhough the optimal situation would have been separate control groups in the proteome analysis.

In this clinical study we identified 486 of the 897 proteins from the porcine study of EBC [16]. The protein concentration was markedly higher in the porcine EBC samples (mean protein concentration 3.08 μg ml^{-1}), possibly due to both inter-species variation and sampling method as the pigs were mechanically ventilated during EBC collection. A study of EBC collected from mechanically ventilated patients could clarify the impact of the collection method.

We identified a markedly higher number of proteins in the EBC samples compared with previous human studies of exhaled proteins [14, 17, 41, 42], although a study of Ma *et al* found 1254 proteins [43]. The EBC proteome in our study has a relatively high overlap in identified proteins with previous studies and the 928 proteins we identified in this study included 107 out of 145 proteins (74%) from the study by Lacombe *et al* [17], 19 out of 44 proteins (43%) from the study by Fumagalli *et al* [14], 110 out of 167 proteins (66%) in the study by Muccilli *et al* [41], 47 out of 257 proteins (18%) in the study by Sun *et al* [42], using TMT labelling and 188 out of 1254 proteins (15%) in the study by Ma *et al* [43]. Various strategies have been used including protein separation by polyacrylamide gel electrophoresis, TMT labelling and data independent acquisition. Ma *et al* applied a data independent acquisition LC-MS/MS analysis strategy which is more sensitive to identify low-abundant proteins compared to the data-dependent acquisition used in

the present study [44]. The findings indicate that EBC proteomics studies can benefit from data independent strategies, although the studies may be difficult to compare since they vary with respect to data filtration and amount of 'missing values'. However, albumin was identified in the EBC in all studies but Fumagallis, and another 53 proteins were identified in our study and at least two of the other studies.

The search for new diagnostic markers of PE has been centred around plasma proteins [40, 45–48]. When using a new biological sample type as EBC, it is reassuring to identify some of the same potential markers as seen in plasma studies. We identified in EBC many proteins related to the same biological processes and cellular responses as observed in the study by Han *et al*, who also observed diagnostic potential of both serotransferrin (TF) and IGHA1 as seen in the porcine study [16, 45].

The lack of standardization in all aspects of EBC research is the most important limitation of our study. Based on protein concentration measurements in a few of the study samples, we knew that the protein concentration was at the detection limit. We discussed whether to pool EBC samples from each of the groups due to the low protein concentration, but ultimately settled on performing single sample proteomic analysis, to have single sample protein data. However, the technical difficulty and limited sensitivity to detect proteins at this concentration resulted in many missing values as in other studies. The investigation of exhaled proteome is in its very early days and even the method to collect the EBC is not standardized. In addition, while successful in characterizing the EBC proteome and identify known and novel PE biomarkers, the applied MS methods could likely be further optimized for high sensitivity, e.g. by utilizing DIA. Mask-based collection of EBC even with real-time biosensors or filter based caption of exhaled proteins might be part of the solution in future studies, but much more research is needed before EBC analysis can transfer to clinical use [49].

In conclusion, we successfully analysed 77 single EBC samples from patients with suspected PE and controls using LFQ nLC MS/MS. The protein concentration in the EBC samples was low, but we identified biologically relevant proteins in the samples including potential biomarkers of pulmonary embolism worth pursuing in follow-up validation studies.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

The authors wishes to thank Professor Søren Risom Kristensen and Professor Marianne Tang Severinsen for fruitful discussions about the study design, data presentation and – interpretation. We thank Mona Brit Hansen for excellent laboratory work.

Ethical statement

Written patient information plus written and oral consent to participate was obtained before EBC collection began. The study was conducted according to the principles embodies in the Helsinki Declaration and approved by the North Denmark Regional Committee on Health Research Ethics (Journal number N-20180086) and registered at ClinicalTrials.gov (Identifier NCT04010760).

Author contributions

I L G designed the study, included patients, collected clinical data, analysed clinical and proteomic data, acquired funding for the study and wrote the original manuscript draft. S J R reviewed and analysed the clinical data, contributed to manuscript draft and visualisation of results. T S V contributed to patient recruitment and reviewed the manuscript. R F B contributed to proteome data analysis and reviewed the manuscript. T B B contributed to proteome data analysis and visualization and writing of the original manuscript. B H acquired funding for the study, conducted the mass spectrometry analysis, analysed the proteome data, contributed to the manuscript draft, review and edition. The final version of the manuscript has been approved for submission to Journal of Breath Research by all authors.

Funding

This study was funded by grants from Heinrich Kopps grant (2018) and the Novo Nordisk Foundation, Proof of Concept Program 2018, application reference 0053310, project title: 'Development of a non-invasive diagnostic kit for pulmonary embolism by use of biomarkers in exhaled breath condensate'. The Mass spectrometry platform was kindly donated by A.P. Møller og Hustru Chastine Mc-Kinney Møllers Fond til almene Formaal.

Conflict of interest

Aalborg University Hospital, Aarhus University and Aalborg University have filed a patent application based on the findings in the porcine study, [16] (WO2020245200 BIOMARKERS FOR PULMONARY EMBOLISM IN EXHALED BREATH CONDENSATE (wipo.int) accessed on 8 May 2023).

I L G, B H and Søren Risom Kristensen are the inventors. S J R, T S V, T B B and R F B have no conflicting interests.

ORCID iDs

Inger Lise Gade https://orcid.org/0000-0003-0533-5531
Tue Bjerg Bennike https://orcid.org/0000-0003-4354-1731

References

[1] Stein P D *et al* 2007 Clinical characteristics of patients with acute pulmonary embolism: data from PIOPED II *Am. J. Med.* **120** 871–9
[2] Prandoni P *et al* 2016 Prevalence of pulmonary embolism among patients hospitalized for syncope *New Engl. J. Med.* **375** 1524–31
[3] Prandoni P *et al* 2023 Prevalence of pulmonary embolism among patients with recent onset of dyspnea on exertion. A cross-sectional study *J. Thromb. Haemost.* **21** 68–75
[4] Tavora F, Crowder C, Kutys R and Burke A 2008 Discrepancies in initial death certificate diagnoses in sudden unexpected out-of-hospital deaths: the role of cardiovascular autopsy *Cardiovasc. Pathol.* **17** 178–82
[5] Tavora F, Crowder C, Sun C and Burke A 2008 Discrepancies between clinical and autopsy diagnoses: a comparison of university, community, and private autopsy practices *Am. J. Clin. Pathol.* **129** 102–9
[6] Konstantinides S V *et al* 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism *Eur. Heart J.* **35** 3033–69
[7] Suzuki T *et al* 2016 Editor's choice-biomarkers of acute cardiovascular and pulmonary diseases *Eur. Heart J. Acute Cardiovasc. Care* **5** 416–33
[8] Giannitsis E and Katus H A 2016 Biomarkers for clinical decision-making in the management of pulmonary embolism *Clin. Chem.* **63** 91–100
[9] Fens N, Douma R A, Sterk P J and Kamphuisen P W 2010 Breathomics as a diagnostic tool for pulmonary embolism *J. Thromb. Haemost.* **8** 2831–3
[10] von C *et al* 2016 Urine proteome analysis as a discovery tool in patients with deep vein thrombosis and pulmonary embolism *Proteomics Clin. Appl.* **10** 574–84
[11] Beck O, Olin A-C and Mirgorodskaya E 2016 Potential of mass spectrometry in developing clinical laboratory biomarkers of nonvolatiles in exhaled breath *Clin. Chem.* **62** 84–91
[12] Beale D J, Jones O A H, Karpe A V, Dayalan S, Oh D, Kouremenos K, Ahmed W and Palombo E 2016 A review of analytical techniques and their application in disease diagnosis in breathomics and salivaomics research *Int. J. Mol. Sci.* **18** 24
[13] Harshman S, Grigsby C and Ott D 2014 Exhaled breath condensate for proteomic biomarker discovery *Chromatography* **1** 108–19
[14] Fumagalli M *et al* 2012 Profiling the proteome of exhaled breath condensate in healthy smokers and COPD patients by LC-MS/MS *Int. J. Mol. Sci.* **13** 13894–910
[15] Gade I L, Schultz J G, Cehofski L J, Kjærgaard B, Severinsen M T, Rasmussen B S, Vorum H, Honoré B and Kristensen S R 2021 Exhaled breath condensate in acute pulmonary embolism; a porcine study of effect of condensing temperature and feasibility of protein analysis by mass spectrometry *J. Breath Res.* **15** 26005
[16] Gade I L, Schultz J G, Brøndum R F, Kjærgaard B, Nielsen-Kudsk J E, Andersen A, Kristensen S R and Honoré B 2021 Putative biomarkers for acute pulmonary embolism in exhaled breath condensate *J. Clin. Med.* **10** 21
[17] Lacombe M, Marie-Desvergne C and Combes F Kraut A, Bruley C, Vandenbrouck Y, Mossuz V C, Couté Y and Brun V 2017 Proteomic characterization of human exhaled breath condensate *J. Breath Res.* **12** 021001
[18] Hüttmann E-M *et al* 2016 Correction: comparison of two devices and two breathing patterns for exhaled breath condensate sampling *PLoS One* **11** e0152620
[19] Korovesi M D I *et al* 2011 Exhaled breath condensate in mechanically ventilated brain-injured patients with no lung injury or sepsis *Anesthesiology* **114** 1118–29
[20] Aliberti S, Morlacchi L C, Faverio P, Fernandez-Botran R, Cosentini R, Mantero M, Peyrani P, Ramirez J, Bordon J and Blasi F 2016 Serum and exhaled breath condensate inflammatory cytokines in community-acquired pneumonia: a prospective cohort study *Pneumonia* **8** 1
[21] Bach J-P *et al* 2015 Measuring compounds in exhaled air to detect alzheimer's disease and parkinson's disease *PLoS One* **10** e0132227
[22] Bloemen K, Lissens G, Desager K and Schoeters G 2007 Determinants of variability of protein content, volume and pH of exhaled breath condensate *Respir. Med.* **101** 1331–7
[23] Czebe K, Barta I, Antus B, Valyon M, Horváth I and Kullmann T 2008 Influence of condensing equipment and temperature on exhaled breath condensate pH, total protein and leukotriene concentrations *Respir. Med.* **102** 720–5
[24] Bhimji A *et al* 2018 Aspergillus galactomannan detection in exhaled breath condensate compared to bronchoalveolar lavage fluid for the diagnosis of invasive aspergillosis in immunocompromised patients *Clin. Microbiol. Infect.* **24** 640–5
[25] Wiśniewski J R and Gaugaz F Z 2015 Fast and sensitive total protein and peptide assays for proteomic analysis *Anal. Chem.* **87** 4110–6
[26] Ludvigsen M, Thorlacius-Ussing L, Vorum H, Moyer M P, Stender M T, Thorlacius-Ussing O and Honoré B 2020 Proteomic characterization of colorectal cancer cells versus normal-derived colon mucosa cells: approaching identification of novel diagnostic protein biomarkers in colorectal cancer *Int. J. Mol. Sci.* **21** 10
[27] Tyanova S, Temu T and Cox J 2016 The MaxQuant computational platform for mass spectrometry-based shotgun proteomics *Nat. Protocols* **11** 2301–19
[28] Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein M Y, Geiger T, Mann M and Cox J 2016 The Perseus computational platform for comprehensive analysis of (prote)omics data *Nat. Methods* **13** 731–40
[29] R Core Team 2022 A language and environment for statistical computing (available at: www.r-project.org/)
[30] RStudio Team 2020 RStudio: integrated development for R (RStudio) (available at: www.rstudio.com/)
[31] Lazar C and Burger T 2022 ImputeLCMD: a collection of methods for left-censored missing data imputation (available at: https://cran.r-project.org/web/packages/imputeLCMD/imputeLCMD.pdf)
[32] Teufel F, Almagro Armenteros J J, Johansen A R, Gíslason M H, Pihl S I, Tsirigos K D, Winther O, Brunak S, von Heijne G and Nielsen H 2022 SignalP 6.0 predicts all five types of signal peptides using protein language models *Nat. Biotechnol.* **40** 1023–5
[33] Bates D, Mächler M, Bolker B and Walker S 2015 Fitting linear mixed-effects models using lme4 *J. Stat. Softw.* **67** 1–48
[34] Szklarczyk D *et al* 2019 STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets *Nucleic Acids Res.* **47** D607–13
[35] Franceschini A *et al* 2013 STRING v9.1: protein-protein interaction networks, with increased coverage and integration *Nucleic Acids Res.* **41** D808–15
[36] Madsen J, Tornøe I, Nielsen O, Koch C, Steinhilber W and Holmskov U 2003 Expression and localization of lung surfactant protein a in human tissues *Am. J. Respir. Cell Mol. Biol.* **29** 591–7

[37] Uhlén M et al 2015 Tissue-based map of the human proteome Science 347 1260419
[38] Watson L M et al 2003 Overexpression of the 78-kDa glucose-regulated protein/immunoglobulin-binding protein (GRP78/BiP) inhibits tissue factor procoagulant activity J. Biol. Chem. 278 17438–47
[39] Luo G, Wang R, Zhou H and Liu X 2021 ALDOA protects cardiomyocytes against H/R-induced apoptosis and oxidative stress by regulating the VEGF/Notch 1/Jagged 1 pathway Mol. Cell. Biochem. 476 775–83
[40] Bruzelius M et al 2016 PDGFB, a new candidate plasma biomarker for venous thromboembolism: results from the VEREMA affinity proteomics study Blood 128 e59–e66
[41] Muccilli V, Saletti R, Cunsolo V, Ho J, Gili E, Conte E, Sichili S, Vancheri C and Foti S 2015 Protein profile of exhaled breath condensate determined by high resolution mass spectrometry J. Pharm. Biomed. Anal. 105 134–49
[42] Sun C, Zhou T, Xie G, Fu S, Gao L, Liao J, Wu Y and Wang G 2019 Proteomics of exhaled breath condensate in stable COPD and non-COPD controls using tandem mass tags (TMTs) quantitative mass spectrometry: a pilot study J. Proteomics 206 103392
[43] Ma L, Xiu G, Muscat J, Sinha R, Sun D and Xiu G 2022 Comparative proteomic analysis of exhaled breath condensate between lung adenocarcinoma and CT-detected benign pulmonary nodule patients Cancer Biomarkers 34 163–74
[44] Bennike T B et al 2018 A cost-effective high-throughput plasma and serum proteomics workflow enables mapping of the molecular impact of total pancreatectomy with islet autotransplantation J. Proteome Res. 17 1983–92
[45] Han B et al 2021 Discovery of plasma biomarkers with data-independent acquisition mass spectrometry and antibody microarray for diagnosis and risk stratification of pulmonary embolism J. Thromb. Haemost. 19 1738–51
[46] Zhang Y-X, Li J-F, Yang Y-H, Huang K, Miao R, Zhai Z-G and Wang C 2018 Identification of haptoglobin as a potential diagnostic biomarker of acute pulmonary embolism Blood Coagul. Fibrinolysis 29 275–81
[47] Insenser M, Montes-Nieto R, Martínez-García M Á, Durán E F, Santiuste C, Gómez V, Kline J A, Escobar-Morreale H F and Jiménez D 2014 Identification of reduced circulating haptoglobin concentration as a biomarker of the severity of pulmonary embolism: a nontargeted proteomic study PLoS One 9 e100902
[48] Ten Cate V et al 2021 Protein expression profiling suggests relevance of noncanonical pathways in isolated pulmonary embolism Blood 137 2681–93
[49] Oberbeckmann-Winter N et al 2022 Exhaled breath condensate as bioanalyte: from collection considerations to biomarker sensing Anal. Bioanal. Chem. 414 1–8