Proteomic characterization of human exhaled breath condensate

Maud Lacombe1,2,3, Caroline Marie-Desvergne1,4, Florence Combes1,2,3, Alexandra Kraut1,2,4, Christophe Bruley1,2,3, Yves Vandenbrouck1,2,3, Véronique Chamel Mossuz1,4, Yohann Couté1,2,3 and Virginie Brun1,2,3

1 Université Grenoble-Alpes, F-38000, Grenoble, France
2 CEA, BIG, Biologie à Grande Echelle, F-38054, Grenoble, France
3 Inserm, Unité 1038, F-38054, Grenoble, France
4 CEA, NanoSafety Platform, Medical Biology Laboratory, F-38054, Grenoble, France

E-mail: virginie.brun@cea.fr

Keywords: exhaled breath condensate, proteomics, mass spectrometry

Abstract

To improve biomedical knowledge and to support biomarker discovery studies, it is essential to establish comprehensive proteome maps for human tissues and biofluids, and to make them publicly accessible. In this study, we performed an in-depth proteomics characterization of exhaled breath condensate (EBC), a sample obtained non-invasively by condensation of exhaled air that contains submicron droplets of airway lining fluid. Two pooled samples of EBC, each obtained from 10 healthy donors, were processed using a straightforward protocol based on sample lyophilization, in-gel digestion and liquid chromatography tandem-mass spectrometry analysis. Two ‘technical’ control samples were processed in parallel to the pooled samples to correct for exogenous protein contamination. A total of 229 unique proteins were identified in EBC among which 153 proteins were detected in both EBC pooled samples. A detailed bioinformatics analysis of these 153 proteins showed that most of the proteins identified corresponded to proteins secreted in the respiratory tract (lung, bronchi). Eight proteins were salivary proteins. Our dataset is described and has been made accessible through the ProteomeXchange database (dataset identifier: PXD007591) and is expected to be useful for future MS-based biomarker studies using EBC as the diagnostic specimen.

Abbreviations

COPD Chronic obstructive pulmonary disease
EBC Exhaled breath condensate
FDR False discovery rate
GO Gene ontology
IBAQ Intensity-based absolute quantification
LC-MS/MS Liquid chromatography tandem-mass spectrometry
PEX Exhaled air endogenous particles
SELDI Surface-enhanced laser desorption/ionization

Introduction

Exhaled breath condensate (EBC) is a biological sample collected by condensing droplets of airway lining fluid present in the exhaled air. It is a highly diluted matrix containing diverse components including salts, phospholipids, metabolites, proteins and inhaled particles such as carbonaceous and metal nanoparticles [1]. EBC collection and physicochemical characterization drive increasing research efforts to explore pathophysiological changes and identify new biomarkers for toxic exposure [2], respiratory diseases [3, 4] and systemic diseases [5]. In this context, the Task Force of the European Respiratory Society has recently published guidelines and recommendations to standardize sample collection and to evaluate technical approaches targeting various analytes in exhaled breath [6]. In the field of
proteomics, a few investigations using surface-enhanced laser desorption/ionization (SELDI) mass spectrometry profiling, two-dimensional electrophoresis and/or liquid chromatography tandem-mass spectrometry analysis (LC-MS/MS) have attempted to characterize the protein content and modifications to EBC in specific situations [7–9]. Some of these studies revealed potential protein biomarkers for asthma [10], chronic obstructive pulmonary disease (COPD) [11, 12] and lung cancer [13, 14]. However, improvements in sampling and analytical procedures are still required to achieve sensitive and comprehensive proteomics characterization of EBC [6, 15].

Although easily collected by a non-invasive technique, EBC is difficult to handle for proteomics analysis as it is extremely diluted (protein concentration < 1 μg mL⁻¹) and because it contains surfactant phospholipids. All previous studies used methods for protein concentration and phospholipid removal, considering them essential for in-depth characterization of the EBC proteome. Moreover, most experiments were performed using pooled EBC samples to improve the detection of low-abundance proteins (near the detection limit in individual samples) and enhance the depth of proteome coverage. In 2012, Bredberg et al. [7] reported the identification of 32 and 116 proteins in exhaled air endogenous particles (PEx) using LC-MS/MS analysis of pooled samples from six and 10 healthy donors, respectively. Exhaled endogenous particles were collected on a specific device and were concentrated using silicon plates before trypsin digestion and LC-MS/MS analysis. Thoroughly, the authors introduced a negative control to correct for non-specific protein identification. In 2015, Mucilli et al. [9] identified 167 proteins in EBC based on LC-MS/MS analysis of a single lyophilized EBC pool collected from nine healthy donors; nine out of the 10 most abundant proteins identified were cytokeratins [9]. More recently, in the context of a lung cancer biomarker discovery study, Lopez-Sanchez et al. [14] collected 49 EBC samples from healthy donors and identified a total of 123 proteins in these EBC specimens based on sample lyophilization, in-solution digestion and LC-MS/MS analysis.

In line with international initiatives that streamline and coordinate efforts in the field of exhaled biomarkers [6], we engaged this study to extend the knowledge of EBC proteome composition and to assess the risk of contamination associated to EBC sample collection and processing. To do this, we performed an in-depth nanoLC-MS/MS analysis of two pooled EBC samples, each of which corresponded to exhalate from 10 healthy donors. Pooled EBC samples were collected using the RTube® commercial device, lyophilized, digested in-gel with trypsin and finally submitted to nanoLC-MS/MS analysis. Based on a rigorous procedure to exclude technical contaminants, 153 unique proteins were reliably identified in both EBC pools.

Materials and methods

EBC collection and preparation
EBC was collected from 20 healthy non-smoking volunteers (seven men and 13 women, mean age: 36 ± 10 years) with no known significant health problems (systemic or respiratory disease) and no symptoms of respiratory tract infection. The RTube® collection device (Respiratory Research Inc., USA) was used to collect EBC samples as previously described [2]. Volunteers breathed normally into the pre-cooled (−20 °C) device for 15 min, using a nose clip to prevent nasal inhalation and exhalation. For each volunteer, the collected sample corresponded to 120 l of exhaled breath condensed in a final volume of 1.5–2 ml [16]. Samples were immediately frozen, dried by lyophilization (−47 °C, 9 kPa, 12 h) and stored at −80 °C. During the EBC sampling procedure, gloves and gowns were used to minimize keratin contamination.

SDS-PAGE and in-gel trypsin digestion
Individual samples were combined to constitute two pools of 10 EBC samples each (characteristics of the subjects included in each pool are presented in supplemental table 1 is available online at stacks.iop.org/JBR/12/021001/mmedia). To produce each pool, 25 μl of Laemmli buffer (glycerol, β mercaptoethanol, SDS, bromophenol blue 1%, Tris-Cl pH 6.8) was added to the first dried sample (sample 1) and centrifuged at 800 g, 4 °C, for 1 min. Sample 1 was then pipetted and added to dried sample 2. These steps were repeated until 10 samples had been combined. Proteins from pooled samples were stacked on the top of a precast polyacrylamide gel (NuPAGE™ 4%–12% bis-Tris protein gel, Invitrogen) and revealed by Coomassie blue staining. Gel pieces containing EBC proteins were manually excised and proteins were digested in-gel with trypsin as previously described [17]. Two control samples (distilled water) were included and processed in parallel with the EBC pools as blanks to allow monitoring for protein contamination occurring during the pre-analytical procedure. Peptide digests were resolubilized in 25 μl of 2% acetonitrile, 0.1% formic acid, and 10 μl was injected into the LC-system.

Mass spectrometry-based proteomic analyses
Peptides resulting from trypsin digestion were analyzed by nanoliquid chromatography combined with tandem-mass spectrometry (Ultimate 3000 coupled to LTQ-Orbitrap Velos Pro, Thermo Scientific) using a 120 min gradient, as previously described [18]. RAW files were processed using MaxQuant [19] version 1.5.3.30. Spectra were searched against the SwissProt database (Homo sapiens taxonomy, December 2015 version) and the pig trypsin sequence. Trypsin was chosen as the enzyme and two missed cleavages were
allowed. Precursor mass error tolerances were set at 20 and 4.5 ppm for first and main searches, respectively. Fragment mass error tolerance was set at 0.5 Da. Peptide modifications allowed during the search were: carbamidomethylation (C, fixed), acetyl (Protein N-term, variable) and oxidation (M, variable). Minimum peptide length was set to seven amino acids. Minimum number of peptides, razor + unique peptides and unique peptides were all set to 1. Maximum false discovery rates (FDR)—calculated by employing a reverse database strategy—were set to 0.01 at peptide and protein levels. Intensity-based absolute quantification (iBAQ) values were calculated from MS intensities of unique + razor peptides. Proteins identified in the reverse database and trypsin were discarded from the list of proteins identified. LC-MS/MS data (original raw files) have been deposited to the ProteomeExchange Consortium via the PRIDE partner repository with the dataset identifier: PXD007591 [21].

Data filtering and mining
Protein contamination is a crucial issue when analyzing EBC. Two types of contamination were considered: (i) technical contamination during sample preparation and (ii) biological contamination by saliva during sample collection. To correct for protein contamination during sample processing, ‘technical’ control samples (distilled water) were processed and analyzed alongside the two pooled EBC samples. For each protein identified, a minimum of 100-fold enrichment between the pooled EBC sample and its corresponding ‘technical’ control sample was required for inclusion in the final EBC protein list. Proteins with an enrichment ratio below 100 were considered as technical contaminants. To evaluate contamination of EBC with salivary proteins, the expression pattern of EBC with salivary proteins, the expression pattern as technical contaminants. To evaluate contamination for inclusion in the corresponding enrichment between the pooled EBC sample and its corresponding ‘technical’ control sample. Based on these criteria, we identified a total of 229 unique proteins in the two pooled EBC samples. More precisely, 175 proteins were present in the first pooled EBC sample, 207 in the second sample, and 153 proteins were common to both pools (table 1). The final list of 153 unique proteins identified in the 2 pooled samples was considered as the ‘core’ proteome of EBC (tables 2 and 3).

Importantly, several previous investigations of EBC protein content reported cytokeratins as major constituents of the EBC proteome [9, 23]. However, this group of proteins can also be present due to technical contamination during sample processing. In this study, following filtering, 10 cytokeratins were reliably identified as true components of the EBC proteome. A group of 10 other proteins, however, were identified in both ‘technical’ control samples with an enrichment in EBC samples below the fixed threshold. These proteins were thus considered to be technical contaminants (table 4). Their specific or highly predominant expression in the skin was confirmed using the Human Protein Atlas database [24].

As EBC samples are obtained from air exhaled through the oral cavity, and even though the RTube collection device contained a saliva trap to separate saliva from the exhaled breath, contamination with salivary proteins had to be assessed. Several studies quantified α-amylase activity levels as a means to assess salivary contamination. Alternatively, the EBC proteome can be compared to the salivary proteome, as characterized by Sivadasan et al [25]. However, the origin of proteins identified in both samples is difficult to determine; does it correspond to true overlap or cross-contamination? In this study, we decided to check the expression pattern for each protein of the ‘core’ EBC proteome using the Human Protein Atlas, which was originally developed as an expression dictionary for all protein-coding genes in human tissues and organs [24], the NextProt database [26] and bibliographic information. We sorted the proteins identified into four different groups: (i) proteins specifically expressed in the salivary glands (n = 8), (ii) proteins expressed both in the salivary gland and in other tissues from the respiratory tract (lung, bronchi and nasopharynx) (n = 94), (iii) proteins not expressed in the salivary glands and expressed in the respiratory tract (n = 49) and (iv) 2 proteins expressed in the tongue, esophagus and skin (tables 2 and 3). Interestingly,

Results and discussion

EBC proteome characterization
Two pools of 10 individual EBC samples and two ‘technical’ control samples were constituted to allow in-depth and reliable characterization of the EBC proteome. Samples were processed as follows: lyophilization, protein concentration using a stacking gel, in-gel digestion with trypsin and analysis of peptide digests by single-shot nanoLC-MS/MS (figure 1). Data processing using 1 significant peptide per protein and a FDR below 1% at the peptide and protein levels led to the identification of 430 proteins in the four samples (supplemental table 2). To extract the ‘core’ EBC proteome, data were further filtered using more stringent criteria: (i) identification with a minimum of two significant peptides per protein, (ii) minimal iBAQ enrichment of 100-fold between each pooled EBC sample and its corresponding ‘technical’ control sample. Based on these criteria, we identified a total of 229 unique proteins in the two pooled EBC samples. More precisely, 175 proteins were present in the first pooled EBC sample, 207 in the second sample, and 153 proteins were common to both pools (table 1). The final list of 153 unique proteins identified in the 2 pooled samples was considered as the ‘core’ proteome of EBC (tables 2 and 3).

Importantly, several previous investigations of EBC protein content reported cytokeratins as major constituents of the EBC proteome [9, 23]. However, this group of proteins can also be present due to technical contamination during sample processing. In this study, following filtering, 10 cytokeratins were reliably identified as true components of the EBC proteome. A group of 10 other proteins, however, were identified in both ‘technical’ control samples with an enrichment in EBC samples below the fixed threshold. These proteins were thus considered to be technical contaminants (table 4). Their specific or highly predominant expression in the skin was confirmed using the Human Protein Atlas database [24].

As EBC samples are obtained from air exhaled through the oral cavity, and even though the RTube collection device contained a saliva trap to separate saliva from the exhaled breath, contamination with salivary proteins had to be assessed. Several studies quantified α-amylase activity levels as a means to assess salivary contamination. Alternatively, the EBC proteome can be compared to the salivary proteome, as characterized by Sivadasan et al [25]. However, the origin of proteins identified in both samples is difficult to determine; does it correspond to true overlap or cross-contamination? In this study, we decided to check the expression pattern for each protein of the ‘core’ EBC proteome using the Human Protein Atlas, which was originally developed as an expression dictionary for all protein-coding genes in human tissues and organs [24], the NextProt database [26] and bibliographic information. We sorted the proteins identified into four different groups: (i) proteins specifically expressed in the salivary glands (n = 8), (ii) proteins expressed both in the salivary gland and in other tissues from the respiratory tract (lung, bronchi and nasopharynx) (n = 94), (iii) proteins not expressed in the salivary glands and expressed in the respiratory tract (n = 49) and (iv) 2 proteins expressed in the tongue, esophagus and skin (tables 2 and 3). Interestingly,
among the 49 proteins expressed in the respiratory tract only, some are mainly expressed in the upper respiratory parts such as serpin B3 (bronchi, naso-pharynx); others are more abundant in the deep lung such as fatty acid-binding protein 5, which is strongly expressed in lung macrophages. At last, some proteins are expressed all along the respiratory tract, such as cystatin-A. While the precise contribution of each respiratory compartment to the EBC content is still under discussion [13], our results bring additional confirmation that EBC may be representative of all the levels of the respiratory tract including deep lung which is a critical target for different toxicants such as nanoparticles.

Functional annotation of the EBC proteome
The list of 145 proteins identified in the two pooled EBC samples (excluding the eight salivary proteins) was submitted to GO-term enrichment analysis [22] to determine functions that were significantly enriched in our EBC proteomic dataset compared to the lung proteome (corresponding to 5469 genes extracted from the Human Protein Atlas). According to this analysis, the main biological processes that were found over-represented in EBC compared to lung were immune system processes, exocytosis and NAD/NADH metabolism (figure 2(A)). Hence, the EBC proteome was found to contain several proteins of the airway mucus including mucin 5B, DMBT1 (deleted

Table 1. Proteins identified in EBC pooled samples using nanoLC-MS/MS analysis.

| Number of significant peptides considered for protein identification | Number of proteins identified in the first EBC pool | Number of proteins identified in the second EBC pool | Total number of identified proteins in EBC | Number of proteins common to both EBC pools |
|---------------------------------------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| 1 peptide                                                   | 267                             | 305                             | 349                             | 188                             |
| 2 peptides                                                  | 175                             | 207                             | 229                             | 153                             |

Figure 1. Workflow for EBC sample pooling, preparation and LC-MS/MS analysis.
Table 2. List of the 145 unique proteins (excluding the eight salivary proteins) identified in both pooled EBC samples by LC-MS/MS with at least two significant peptides (FDR 1%).

| Protein number | Protein accession number (UniProt) | Protein name | Number of peptides (razor + unique) | Salivary glands + respiratory tract | Tongue, esophagus and skin | Respiratory tract only |
|----------------|-----------------------------------|--------------|-------------------------------------|------------------------------------|----------------------------|----------------------|
| 1              | P15924                            | Desmoplakin  | 69                                  | x                                  |                            |                      |
| 2              | P02538                            | Keratin, type II cytoskeletal 6A | 53                                  |                                    | x                          |                      |
| 3              | P02768                            | Serum albumin | 44                                  |                                    |                            | x                    |
| 4              | P08779                            | Keratin, type I cytoskeletal 16 | 29                                  |                                    |                            | x                    |
| 5              | Q02413                            | Desmoglein-1 | 24                                  |                                    |                            | x                    |
| 6              | P07355                            | Annexin A2; putative annexin A2-like protein | 22                                  |                                    |                            | x                    |
| 7              | P14923                            | Junction plakoglobin | 22                                  |                                    |                            | x                    |
| 8              | P02788                            | Lactotransferrin | 21                                  |                                    |                            | x                    |
| 9              | Q9HC84                            | Mucin 5B     | 21                                  |                                    |                            | x                    |
| 10             | P29508                            | Serpin B3    | 20                                  |                                    |                            |                      |
| 11             | P63261                            | Actin, cytoplasmic 2 | 19                                  |                                    |                            | x                    |
| 12             | Q8N1N4                            | Keratin, type II cytoskeletal 78 | 18                                  |                                    |                            | x                    |
| 13             | Q04695                            | Keratin, type I cytoskeletal 17 | 18                                  |                                    |                            | x                    |
| 14             | P01876                            | Ig alpha-1 chain C region | 16                                  |                                    |                            | x                    |
| 15             | Q01469                            | Fatty acid-binding protein 5, epidermal | 15                                  |                                    |                            | x                    |
| 16             | P31944                            | Caspase-14   | 15                                  |                                    |                            |                      |
| 17             | P01833                            | Polymeric immunoglobulin receptor | 15                                  |                                    |                            | x                    |
| 18             | P06733                            | Alpha-enolase | 15                                  |                                    |                            | x                    |
| 19             | P25311                            | Zinc-alpha-2-glycoprotein | 15                                  |                                    |                            | x                    |
| 20             | Q15149                            | Plectin      | 15                                  |                                    |                            | x                    |
| 21             | P19013                            | Keratin, type II cytoskeletal 4 | 13                                  |                                    |                            | x                    |
| 22             | Q6KB66                            | Keratin, type II cytoskeletal 80 | 13                                  |                                    |                            | x                    |
| 23             | Q08188                            | Protein-glutamine gamma-glutamyltransferase E | 12                                  |                                    |                            |                      |
| 24             | P13646                            | Keratin, type I cytoskeletal 13 | 11                                  |                                    |                            |                      |
| 25             | Q86YZ3                            | Hornerin     | 11                                  |                                    |                            | x                    |
| 26             | P04259                            | Keratin, type II cytoskeletal 8R | 10                                  |                                    |                            | x                    |
| 27             | P02545                            | Prelamin-A/Clr lamin-A/C | 10                                  |                                    |                            | x                    |
| 28             | P04083                            | Annexin A1   | 10                                  |                                    |                            | x                    |
| 29             | P11021                            | 78 kDa glucose-regulated protein | 10                                  |                                    |                            | x                    |
| 30             | P02787                            | Serotransferrin | 9                                   |                                    |                            | x                    |
| 31             | P04040                            | Catalase     | 9                                   |                                    |                            | x                    |
| 32             | P31151                            | Protein S100-A7 | 9                                   |                                    |                            | x                    |
| 33             | P31947                            | 14-3-3 protein sigma | 9                                   |                                    |                            | x                    |
| Protein number | Protein accession number | Protein name                  | Number of peptides | Salivary glands + respiratory tract | Tongue, esophagus and skin | Respiratory tract only |
|----------------|--------------------------|-------------------------------|--------------------|-------------------------------------|-----------------------------|------------------------|
| 34             | Q96P63                   | Serpin B12                    | 9                  | x                                   |                             |                        |
| 35             | P14618                   | Pyruvate kinase PKM           | 9                  | x                                   |                             |                        |
| 36             | P60174                   | Triosephosphate isomerase     | 9                  | x                                   |                             |                        |
| 37             | Q06830                   | Peroxiredoxin-1               | 9                  | x                                   |                             |                        |
| 38             | P01040                   | Cystatin-A                    | 8                  | x                                   |                             |                        |
| 39             | P05089                   | Arginase-1                    | 8                  | x                                   |                             |                        |
| 40             | P01834                   | Ig kappa chain C region       | 8                  | x                                   |                             |                        |
| 41             | P04406                   | Glyceraldehyde-3-phosphate dehydrogenase | 8 | x                                   |                             |                        |
| 42             | P0DMV9                   | Heat shock 70 kDa protein 1B  | 8                  | x                                   |                             |                        |
| 43             | P13639                   | Elongation factor 2           | 8                  | x                                   |                             |                        |
| 44             | P35579                   | Myosin-9                      | 8                  | x                                   |                             |                        |
| 45             | P68371                   | Tubulin beta-4B chain         | 8                  | x                                   |                             |                        |
| 46             | Q8WVV4                   | Protein POF1B                 | 8                  | x                                   |                             |                        |
| 47             | O75635                   | Serpin B7                     | 7                  | x                                   |                             |                        |
| 48             | P01857                   | Ig gamma-1 chain C region     | 7                  | x                                   |                             |                        |
| 49             | P61626                   | Lysozyme-C                    | 7                  | x                                   |                             |                        |
| 50             | P68363                   | Tubulin alpha-1B chain        | 7                  | x                                   |                             |                        |
| 51             | P01009                   | Alpha-1-antitrypsin; short peptide from AAT | 6 | x                                   |                             |                        |
| 52             | P07900                   | Heat shock protein HSP 90-alpha| 6                  | x                                   |                             |                        |
| 53             | Q9NZIH8                  | Interleukin-36 gamma          | 6                  | x                                   |                             |                        |
| 54             | O43707                   | Alpha-actinin-4; alpha-actinin-1 | 6     | x                                   |                             |                        |
| 55             | O75223                   | Gamma-glutamylcyclotransferase | 6                  | x                                   |                             |                        |
| 56             | P00338                   | L-lactate dehydrogenase A chain | 6             | x                                   |                             |                        |
| 57             | P07339                   | Cathepsin D                   | 6                  | x                                   |                             |                        |
| 58             | P62987                   | Ubiquitin-60S ribosomal protein L40 | 6 | x                                   |                             |                        |
| 59             | P10599                   | Thioredoxin                   | 6                  | x                                   |                             |                        |
| 60             | Q9UGM3                   | Deleted in malignant brain tumors 1 protein | 6 | x                                   |                             |                        |
| 61             | Q9UI42                   | Carboxypeptidase A4           | 6                  | x                                   |                             |                        |
| 62             | P47929                   | Galectin-7                    | 5                  | x                                   |                             |                        |
| 63             | Q13867                   | Bleomycin hydrolase           | 5                  | x                                   |                             |                        |
| 64             | Q6P4A8                   | Phospholipase B-like I        | 5                  | x                                   |                             |                        |
| 65             | O75369                   | Filamin-B                     | 5                  | x                                   |                             |                        |
| 66             | P00441                   | Superoxide dismutase [Cu-Zn]  | 5                  | x                                   |                             |                        |
| Protein number | Protein accession number (UniProt) | Protein name | Number of peptides (razor + unique) | Expression pattern$^a$ |
|----------------|-----------------------------------|--------------|-------------------------------------|----------------------|
| 67             | P04792                            | Heat shock protein beta-1 | 5                                  | x                    |
| 68             | P11142                            | Heat shock cognate 71 kDa protein | 5                                  | x                    |
| 69             | P58107                            | Epiplakin     | 5                                  | x                    |
| 70             | P60842                            | Eukaryotic initiation factor 4A-1 | 5                                  | x                    |
| 71             | P62937                            | Peptidyl-prolyl cis-trans isomerase A | 5                                  | x                    |
| 72             | P63104                            | 14-3-3 protein zeta/delta | 5                                  | x                    |
| 73             | Q92820                            | Gamma-glutamyl hydrolase | 5                                  | x                    |
| 74             | O75342                            | Arachidonate 12-lipoxygenase, 12R-type | 4                                  | x                    |
| 75             | P09211                            | Glutathione S-transferase P | 4                                  | x                    |
| 76             | P31025                            | Lipocalin-1   | 4                                  | x                    |
| 77             | P48594                            | Serpin B4     | 4                                  | x                    |
| 78             | Q14574                            | Desmocollin-3 | 4                                  | x                    |
| 79             | Q5T750                            | Skin-specific protein 32 | 4                                  | x                    |
| 80             | Q6UWP8                            | Suprabasin    | 4                                  | x                    |
| 81             | O6O911                            | Cathepsin L2  | 4                                  | x                    |
| 82             | P00558                            | Phosphoglycerate kinase 1 | 4                                  | x                    |
| 83             | P04075                            | Fructose-bisphosphate aldolase A | 4                                  | x                    |
| 84             | P07384                            | Calpain-1 catalytic subunit | 4                                  | x                    |
| 85             | P0CG05                            | Ig lambda-2 chain C regions | 4                                  | x                    |
| 86             | P18206                            | Vinculin      | 4                                  | x                    |
| 87             | P62258                            | 14-3-3 protein epsilon | 4                                  | x                    |
| 88             | P68871                            | Hemoglobin subunit beta | 4                                  | x                    |
| 89             | Q9C075                            | Keratin, type I cytoskeletal 23 | 4                                  | x                    |
| 90             | A8K2U0                            | Alpha-2-macroglobulin-like protein 1 | 3                                  | x                    |
| 91             | P00738                            | Haptoglobin   | 3                                  | x                    |
| 92             | P01011                            | Alpha-1-antichymotrypsin | 3                                  | x                    |
| 93             | P02763                            | Alpha-1-acid glycoprotein 1 | 3                                  | x                    |
| 94             | P18510                            | Interleukin-1 receptor antagonist protein | 3                                  | x                    |
| 95             | P22528                            | Cornifin-B    | 3                                  | x                    |
| 96             | P30740                            | Leukocyte elastase inhibitor | 3                                  | x                    |
| 97             | P80188                            | Neutrophil gelatinase-associated lipocalin | 3                                  | x                    |
| 98             | Q15828                            | Cystatin-M    | 3                                  | x                    |
| 99             | Q9HCY8                            | Protein S100-A14 | 3                                  | x                    |
| Protein number | Protein accession number (UniProt) | Protein name | Number of peptides (razor + unique) | Salivary glands + respiratory tract | Tongue, esophagus and skin | Respiratory tract only |
|----------------|-----------------------------------|--------------|-------------------------------------|-----------------------------------|---------------------------|------------------------|
| 100            | P01623                            | Ig kappa chain V-III region | 3                                   | x                                 | x                         |                        |
| 101            | P01877                            | Ig alpha-2 chain C region | 3                                   | x                                 | x                         |                        |
| 102            | P06396                            | Gelsolin      | 3                                   | x                                 | x                         |                        |
| 103            | P14735                            | Insulin-degrading enzyme | 3                                   | x                                 | x                         |                        |
| 104            | P20933                            | N(4)-(beta-N-acetylglucosaminy) - L-asparaginase | 3                                   | x                                 | x                         |                        |
| 105            | P25788                            | Proteasome subunit alpha type-3 | 3                                   | x                                 | x                         |                        |
| 106            | P26641                            | Elongation factor 1-gamma | 3                                   | x                                 | x                         |                        |
| 107            | P36952                            | Serpin B5     | 3                                   | x                                 | x                         |                        |
| 108            | P40926                            | Malate dehydrogenase, mitochondrial | 3                                   | x                                 | x                         | x                      |
| 109            | Q9Y6R7                            | IgGFC-binding protein | 3                                   | x                                 | x                         |                        |
| 110            | O95274                            | Ly6/PLAUR domain-containing protein 3 | 2                                   | x                                 | x                         |                        |
| 111            | P00491                            | Purine nucleoside phosphorylase | 2                                   | x                                 | x                         |                        |
| 112            | P04080                            | Cystatin-B     | 2                                   | x                                 | x                         |                        |
| 113            | P09972                            | Fructose-bisphosphate aldolase C | 2                                   | x                                 | x                         |                        |
| 114            | P19012                            | Keratin, type 1 cytoskeletal 15 | 2                                   | x                                 | x                         |                        |
| 115            | P20930                            | Filaggrin      | 2                                   | x                                 | x                         |                        |
| 116            | Q96FX8                            | p53 apoptosis effector related to PMP-22 | 2                                   | x                                 | x                         |                        |
| 117            | Q9UIY8                            | Serpin B13    | 2                                   | x                                 | x                         |                        |
| 118            | P01625                            | Ig kappa chain V-IV region Len | 2                                   | x                                 | x                         |                        |
| 119            | P01765                            | Ig heavy chain V-III region TIL | 2                                   | x                                 | x                         |                        |
| 120            | P01766                            | Ig heavy chain V-III region BRO | 2                                   | x                                 | x                         |                        |
| 121            | P01860                            | Ig gamma-3 chain C region | 2                                   | x                                 | x                         |                        |
| 122            | P01871                            | Ig mu chain C region | 2                                   | x                                 | x                         |                        |
| 123            | P05090                            | Apolipoprotein D | 2                                   | x                                 | x                         |                        |
| 124            | P06870                            | Kallikrein-1   | 2                                   | x                                 | x                         |                        |
| 125            | P07858                            | Cathepsin B    | 2                                   | x                                 | x                         |                        |
| 126            | P08865                            | 40S ribosomal protein SA | 2                                   | x                                 | x                         |                        |
| 127            | P11279                            | Lysosome-associated membrane glycoprotein 1 | 2                                   | x                                 | x                         |                        |
| 128            | P13473                            | Lysosome-associated membrane glycoprotein 2 | 2                                   | x                                 | x                         |                        |
| 129            | P19971                            | Thymidine phosphorylase | 2                                   | x                                 | x                         |                        |
| 130            | P23284                            | Peptidyl-prolyl cis-trans isomerase B | 2                                   | x                                 | x                         |                        |
| 131            | P23396                            | 40S ribosomal protein S3 | 2                                   | x                                 | x                         |                        |
| 132            | P25705                            | ATP synthase subunit alpha, mitochondrial | 2                                   | x                                 | x                         |                        |
Table 2. (Continued.)

| Protein number | Protein accession number (UniProt) | Protein name                                           | Number of peptides (razor + unique) | Expression patterna |
|----------------|-----------------------------------|--------------------------------------------------------|-------------------------------------|---------------------|
| 133            | P27482                            | Calmodulin-like protein 3                              | 2                                   | x                   |
| 134            | P31949                            | Protein S100-A11                                       | 2                                   | x                   |
| 135            | P40121                            | Macrophage-capping protein                             | 2                                   | x                   |
| 136            | P42357                            | Histidine ammonia-lyase                                | 2                                   | x                   |
| 137            | P47756                            | F-actin-capping protein subunit beta                    | 2                                   | x                   |
| 138            | P48637                            | Glutathione synthetase                                 | 2                                   | x                   |
| 139            | P49720                            | Proteasome subunit beta type-3                         | 2                                   | x                   |
| 140            | P50395                            | Rab GDP dissociation inhibitor beta                     | 2                                   | x                   |
| 141            | P59998                            | Actin-related protein 2/3 complex subunit 4            | 2                                   | x                   |
| 142            | P61160                            | Actin-related protein 2                                | 2                                   | x                   |
| 143            | P61916                            | Epididymal secretory protein E1                        | 2                                   | x                   |
| 144            | P63244                            | Guanine nucleotide-binding protein subunit beta-2-like 1| 2                                   | x                   |
| 145            | Q9BRQ50                           | Three prime repair exonuclease 2                       | 2                                   | x                   |

Expression pattern for each protein was determined using the Human Protein Atlas (24), the NextProt database (26) and bibliographic information.
in malignant brain tumors 1) protein and alpha-1-antitrypsin [27]. Mucosal secretion prevents adherence of pathogens to the airway epithelial cells and ensures their clearance by the mucociliary escalator, together with inhaled particles. Lysozyme and lactoferrin which are the two most abundant antibacterial proteins secreted into the respiratory tract were also identified in our dataset as well as a myriad of proteins secreted by immune system cells [28]. In general, these results demonstrate that EBC constitutes a relevant matrix to study major physiological functions of the respiratory tract, especially mucosal layer secretion, innate and adaptive antimicrobial defense mechanisms and clearance of inhaled particles [28, 29].

Comparison with previous studies
Our experimental design and the dataset produced (i.e. the list of 153 proteins identified in both pooled EBC samples including the 8 salivary proteins) were compared to the two most extensive EBC proteome maps previously described for healthy subjects [7, 9]. In 2015, Mucilli et al [7] collected EBC from nine non-smoking volunteer donors using a Turbo DECCS device (Medivac, Italy). Samples were pooled to create a single EBC sample with a final volume of 65 ml (equivalent to 1800 l of exhaled breath). After lyophilization, in-gel digestion and LC-MS/MS analysis, these authors identified 167 proteins (two significant peptides per protein, FDR 1%), 77 of which were also included in our protein list (figure 2(B), supplemental table 2). Unlike our procedure, Mucilli et al [7] omitted a control to assess contamination during sample processing, and the eight most abundant proteins in their dataset were cytokeratins, representing 48% of the total emPAI (exponentially modified protein abundance index) [30].

Another proteomics study was performed in 2012 by Bredberg et al [7] to characterize the protein composition of endogenous particles in exhaled air (PEx). These authors used a specific sampling procedure involving silicon plates. Two pooled samples (obtained from six and 10 subjects with forced exhalation) and a negative control (sampling device exposed to ambient air and processed in parallel with the two pooled samples) were analyzed by LC-MS/MS after in-gel digestion. This analysis identified 124 proteins from the two pooled samples, but only 24 proteins were shared by both pools, as a result of the high variability of PEx sample collection. Among the 124 proteins identified in at least one pooled sample, 36 were also identified in our dataset (figure 2(B), supplemental table 2). As already discussed by Mucilli et al [9], these data demonstrate that the sampling method can influence the protein composition of the collected samples. For instance, in 2012, a PEx sampling technique described by Larsson et al [31] was shown to be more efficient in collecting albumin and surfactant protein A than classical EBC collection. Accordingly, no surfactant protein was identified by Mucilli et al [9] and we could detect surfactant protein A in the second EBC pool only (supplemental table 2).

Importantly, 59 proteins from our dataset were identified in neither of these previous studies. A complementary analysis using GO-term annotation [32] showed that these 59 proteins have the same functional distribution between the different proteomic
datasets (Figure 2(C)). All together, these data demonstrate that our analytical procedure did not enrich a specific subproteome but merely extended the coverage of EBC proteome. Undoubtedly, the use of a 2 h LC gradient improved peptide distribution throughout MS/MS analysis and enabled the identification of these novel EBC proteins.

Biomedical potential of EBC proteome
As a non-invasive specimen, EBC could be used for biomarker discovery and analysis. In line with these potential applications, comparative proteomics studies identified biomarker candidates for a variety of pulmonary diseases, including COPD [11, 12], asthma [10], pulmonary emphysema with α-1-antitrypsin deficiency [8] and lung cancer [13, 14]. In agreement with these studies, some of these biomarker candidates (such as α-1-antitrypsin, hornerin, cytokeratins 6A and 6B) were identified in our EBC proteomics dataset. However, our study also identified 10 proteins with high abundance in the two ‘technical’ control samples, including dermcidin, which was recently
selected as a potential biomarker for lung cancer in EBC [14]. The expression pattern for dermcidin may have been modified by tumorigenesis processes (in healthy individuals, dermcidin is not expressed in the respiratory tract), but its presence might also be a technical artefact. This result emphasizes the importance of reliable reference proteome datasets to support clinical biomarker studies [10, 15] and occupational health monitoring of workers exposed to engineered nanoparticles [33].

Most published investigations of the EBC proteome were performed using pooled and lyophilized samples to counteract dilution and favor the detection of low-abundant proteins. However, pooling of EBC samples precludes the evaluation of biological variance which is known to be influenced by age, gender, height and other factors [6, 34]. In our study, we optimized a straightforward analytical procedure based on sample lyophilization, in-gel digestion and nanoLC-MS/MS analysis to characterize EBC specimens. Interestingly, only 40% of each of the peptide digests obtained from 10 healthy subjects was required for injection into the liquid chromatography system before MS/MS analysis. Undoubtedly, this opens the possibility to work with larger sample cohorts, at individual scale using shotgun LC-MS/MS or better still, targeted proteomics approaches.

Recently, shotgun nanoLC-MS/MS experiments were performed at individual scale using EBC samples from 49 healthy volunteers [14]. However, after sample concentration and digestion, very few proteins were identified (an average of 13 proteins per EBC sample) illustrating the difficulty to process sub-microgram protein amounts and to achieve in-depth proteome characterization. In this context, targeted proteomics methods such as selected reaction monitoring (SRM) [35] appear extremely promising. SRM—also referred to as multiple reaction monitoring—is a highly selective MS-based technique that overcomes some limitations of untargeted LC-MS/MS methods. SRM analyses offer the unique possibility to specifically and simultaneously monitor the signatures—so called SRM transitions—of hundreds of preselected peptides generated by protein digestion. Due to its high selectivity, SRM methodology is inherently more sensitive than MS/MS and is especially adapted to the detection of low-abundant proteins in biological matrices. In addition, when combined with isotope-dilution quantification standards, SRM experiments can provide quantitative data for each protein targeted. Likely, proteins identified from untargeted LC-MS/MS analyses of EBC pools will be detectable and quantifiable at individual scale using SRM approaches.

Conclusion

Over the last decade, significant advances in MS-based proteomics instrumentation and methodologies have supported the establishment of comprehensive proteomics maps for human tissues and biofluids. These characterization efforts were sustained by several international research initiatives, such as the Human Proteome Project (HPP) [36–38]. Reliable proteomics surveys, most of which were acquired by LC-MS/MS, are now available for human tissues and biofluids in public repositories. Simultaneously, the European Respiratory Society and the American Thoracic Society have provided recommendations and guidelines to increase the reliability and comparability of exhaled biomarker studies [6]. As a contribution to this field, we performed an in-depth and reliable characterization of the EBC proteome for healthy subjects, taking into account potential exogenous (technical) and endogenous (salivary) sources of protein contaminants. Undoubtedly, this dataset will support future clinical studies dedicated to the discovery of novel protein biomarkers for pulmonary diseases and toxic exposure.

Acknowledgments

We are grateful to Mathilde Louwagie and the team at EDyP for scientific discussions and technical support. We thank Maighread Gallagher–Gambarelli for editing services. This study was supported by grants from the CEA Toxicologie program and the Commissariat à l’Énergie Atomique et aux Energies Alternatives, by the COST Action Clinimark (CA16113) supported by COST (European Cooperation in Science and Technology), by the French National Research Agency in the framework of the ‘Investissements d’avenir’ program (ANR-15-IDEX-02, LIF3 project) and by the ‘Investissement d’Avenir Infrastructures Nationales en Biologie et Santé’ program (ProFI project, ANR-10-INBS-08).

ORCID iDs

Virginie Brun @ https://orcid.org/0000-0002-9098-8707

References

[1] Kuban P and Forest F 2013 Exhaled breath condensate: determination of non-volatile compounds and their potential for clinical diagnosis and monitoring. A review Anal. Chim. Acta 805 1–18
[2] Marie-Desvergne C et al 2016 Assessment of nanoparticles and metal exposure of airport workers using exhaled breath condensate J. Breath Res. 10 036006
[3] Lo Tarn Loï A T et al 2013 Acute and chronic inflammatory responses induced by smoking in individuals susceptible and non-susceptible to development of COPD: from specific disease phenotyping towards novel therapy. Protocol of a cross-sectional study BMJ Open 3 e002178
[4] Santini G et al 2016 Exhaled and non-exhaled non-invasive markers for assessment of respiratory inflammation in patients with stable COPD and healthy smokers J. Breath Res. 10 017102
[5] Alkhouri N et al 2015 Breathprints of childhood obesity: changes in volatile organic compounds in obese children compared with lean controls Pediatr. Obes. 10 23–9
[6] Horvath I et al 2017 A European Respiratory Society technical standard: exhaled biomarkers in lung disease Euro. Respir. J. 49 1609965
[7] Bredberg A et al 2012 Exhaled endogenous particles contain lung proteins Clin. Chem. 58 431–40
[8] Fumagalli M et al 2008 Proteomic analysis of exhaled breath condensate from single patients with pulmonary embolism associated to alpha1-antitrypsin deficiency J. Proteomics 71 211–21
[9] Muccilli V et al 2015 Protein profile of exhaled breath condensate determined by high resolution mass spectrometry J. Pharm. Biomed. Anal. 105 134–49
[10] Bloemen K et al 2011 A new approach to study exhaled proteins as potential biomarkers for asthma Clin. Exp. Allergy 41 346–56
[11] 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
[12] Koczuilla A R et al 2012 Alpha-1 antitrypsin is elevated in exhaled breath condensate and serum in exacerbated COPD patients Respir. Med. 106 120–6
[13] Hayes S A et al 2016 Exhaled breath condensate for lung cancer protein analysis: a review of methods and biomarkers J. Breath Res. 10 034001
[14] Lopez-Sanchez L M et al 2017 Exhaled breath condensate biomarkers for the early diagnosis of lung cancer using proteomics Am. J. Physiol. Lung Cell Mol. Physiol. 313 L664–L676
[15] 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
[16] de Lema I B, Gonzalez M, Vigil L and Casan P 2005 Exhaled breath condensate: standardized collection of samples from healthy volunteers Arch. Bronconeumol. 41 584–6
[17] Casabona M G, Vandembrourght Y, Attree I and Coute Y 2013 Proteomic characterization of Pseudomonas aeruginosa PA01 inner membrane Proteomics 13 2419–23
[18] Millbradt J et al 2014 Proteomic analysis of the multimeric nuclear egress complex of human cytomegalovirus Mol. Cell. Proteomics 13 2132–46
[19] Cox J and Mann M 2008 MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification Nat. Biotechnol. 26 1367–72
[20] Schwanhausser B et al 2011 Global quantification of mammalian gene expression control Nature 473 337–42
[21] Vascaino J A et al 2016 Update of the PRIDE database and its related tools Nucl. Acids Res. 2016 44 11033
[22] Yu G, Wang I G, Han Y and He Q Y 2012 Clusterprofiler: an R package for comparing biological themes among gene clusters Omics. J. Integr. Biol. 16 284–7
[23] Gesner C et al 2008 Presence of cytokeratins in exhaled breath condensate of mechanical ventilated patients Respir. Med. 102 299–306
[24] Uhlen M et al 2010 Towards a knowledge-based human protein atlas Nat. Biotechnol. 28 1246–50
[25] Sivadasan P et al 2015 Human salivary proteome—a resource of potential biomarkers for oral cancer J. Proteomics 127 89–95
[26] Lane L et al 2012 neXtProt: a knowledge platform for human proteins Nucl. Acids Res. 40 D76–83
[27] Joo N S, Evans I A, Cho H J, Park I H, Engelhardt J F and Wine J 2015 Proteomic analysis of pure human airway gland mucus reveals a large component of protective proteins PloS One 10 e1116756
[28] Iwasaki A, Foxman E F and Molony R D 2017 Early local immune defences in the respiratory tract Nat. Rev. Immunol. 17 70–20
[29] Colombo C et al 2011 Analysis of inflammatory and immune response biomarkers in sputum and exhaled breath condensate by a multi-parametric biochip array in cystic fibrosis Int. J. Immunopathol. Pharmacol. 24 4123–32
[30] Ishihama Y et al 2005 Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein Mol. Cell. Proteomics 4 1265–72
[31] Larsson P et al 2012 Surfactant protein A and albumin in particles in exhaled air Respir. Med. 106 197–204
[32] Schmidt A, Forne I and Imhof A 2014 Bioinformatic analysis of proteomics data BMC Syst. Biol. (Suppl. 2) S3
[33] Bergamaschi E 2012 Human biomonitoring of engineered nanoparticles: an appraisal of critical issues and potential biomarkers J. Nanospectr. 2012 564121
[34] 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
[35] Lange V, Piccitti P, Domon B and Aebersold R 2008 Selected reaction monitoring for quantitative proteomics: a tutorial Mol. Syst. Biol. 4 222
[36] Legrain P et al 2011 ‘The human proteome project: current state and future direction Mol. Cell. Proteomics 10 009993
[37] Ommen G S 2017 Advances of the HUPO Human Proteome Project with broad applications for life sciences research Expert. Rev. Proteomics 14 109–11
[38] Carapito C et al 2013 Computational and mass-spectrometry-based workflow for the discovery and validation of missing human proteins: application to chromosomes 2 and 14 J. Proteome Res. 14 3621–34