ISSN 0027-1314, Moscow University Chemistry Bulletin, 2016, Vol. 71, No. 2, pp. 134–139. © Allerton Press, Inc., 2016.
Original Russian Text © K.U. Fedorchenko, A.M. Ryabokon, A.S. Kononikhin, S.I. Mitrofanov, V.V. Barmin, O.V. Pikin, E.H. Anaev, I.V. Gachok, I.A. Popov, E.N. Nikolaev, A.G. Chuchalin, S.D. Varfolomeev, 2016, published in Vestnik Moskovskogo Universiteta. Khimiya, 2016, No. 2, pp. 112–119.

Early Diagnosis of Lung Cancer Based on Proteome Analysis of Exhaled Breath Condensate

K. U. Fedorchenko[a,b], A. M. Ryabokon[a,c], A. S. Kononikhin[a,d,j], S. I. Mitrofanov[b], V. V. Barmin[e], O. V. Pikin[e], E. H. Anaev[f], I. V. Gachok[c], I. A. Popov[a,d], E. N. Nikolaev[a,d], A. G. Chuchalin[f], and S. D. Varfolomeev[a,d]

[a]*Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, Russia*
[b]*International Biotechnological Center, Moscow State University, Moscow, Russia*
[c]*Department of Chemistry, Moscow State University, Moscow, Russia*
[d]*Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, Moscow, Russia*
[e]*Herzen Moscow Oncology Research Institute, Moscow, Russia*
[f]*Research Institute of Pulmonology, Federal Biomedical Agency of Russia, Moscow, Russia*
[j]*Kulakov Research Center for Obstetrics, Gynecology and Perinatology, Ministry of Healthcare of the Russian Federation, Moscow, Russia*
e-mail: amryabokon@gmail.com

Received December 1, 2015

Abstract—A comparative study of the exhaled-breath-condensate (EBC) proteome that was obtained for four donor groups was carried out using ion cyclotron resonance mass spectrometry with electrospray ionization. The groups included subjects with diagnosed lung cancer, chronic obstructive pulmonary disease, community-acquired pneumonia, and healthy nonsmoking control subjects. More than 300 proteins were identified, while 19 of them were found in the EBC samples of the donors who were diagnosed with lung cancer in the early stages and are potentially significant in the development of a diagnostic lung-cancer biomarker panel. It was shown that the EBC protein profiles of different donor groups can be distinguished. It may be possible to highlight a specific protein group that is typical for certain conditions/diseases of the respiratory system. Thus, the EBC analysis could be a promising non-invasive method for early diagnosis of lung cancer.

Keywords: lung cancer, diagnosis, exhaled breath condensate, proteomic analysis, chromatography–mass spectrometry

DOI: 10.3103/S0027131416020036

INTRODUCTION

The respiratory system performs essential functions of life support and reflects human lifestyles and health. Chemical breath tests have a wide range of applications from approved FDA (United States) measurement of exhaled nitric oxide fraction for monitoring the effectiveness of anti-inflammatory therapy in asthma to the determination of volatile organic compounds (VOCs) and profiling of non-volatile biomarkers in a cooled breath sample, which is called exhaled breath condensate (EBC) [1, 2]. The breath test can be easily carried out as it is non-invasive; it allows clinicians and researchers to evaluate the processes that are occurring in the human body. Even in the case of patients who are very ill such sampling can be performed and repeated in short intervals [3]. Thus, the respiration study can be successfully used in screening programs [4].

Along with well-known components such as hydrogen, oxygen, carbon dioxide, inert gases and water vapor, exhaled breath also contains thousands of volatile and non-volatile components, mainly in trace amounts, making their detection a rather complicated challenge. The use of modern highly sensitive technologies in the sample analysis is the basis of the accurate identification of these biomaterials. The use of innovative technologies, such as metabolomics, proteomics, and mass spectrometry, has great potential in biomarker profiling of exhaled breath [5]. Biomarker evaluation of exhaled breath is necessary for understanding disease pathomechanisms, as well as for the prescription of the appropriate therapy [1].

Lung cancer is one of the most lethal types of cancer and is characterized by the highest mortality [6, 7]; the 5-year survival rate is only 15% in lung cancer [8]. If the cancer is diagnosed when it is still localized, a surgical procedure is performed, the prognosis improves, and the 5-year survival reaches 52% [9]. Early cancer diagnosis provides treatment success and a decrease in mortality. However, while screening programs have enabled a reduction in mortality and improvement of the prognosis in bowel, breast, and cervical cancers [10], lung cancer programs are still unsuccessful [11].

Table 1. The characteristics of donor groups

Parameter	Donor group			
	COPD	pneumonia	lung cancer	control
Number of people	17	13	26	23
Age, years	64.7 ± 4.7	36.2 ± 12.2	56.5 ± 11.5	27.5 ± 4.8
Men, heads (%)	13 (76)	7 (54)	21 (81)	10 (43)
Women, heads (%)	4 (24)	6 (46)	5 (19)	13 (57)
Still smoking, former smoker, non-smoker	12/5/0	4/2/7	14/4/8	0/0/23
Stages of the disease	0/3/10/4[a]	2/4/2/3/2[b]	1/8/12/5[c]	[d]
Histological type	–	–	11/4/1/10[e]	–

[a]Stages of COPD: I/II/III/IV, ≥2 positive *Anthonisen* criteria; [b]risk classes according to PSI: I/II/III/IV/V; [c]stages of lung cancer: I/II/III/IV; [d]no symptoms of allergy, chronic respiratory diseases and acute respiratory symptoms during 2 months prior to EBC sampling; [e]squamous cell carcinoma/adenocarcinoma/angiocarcinoma/other types of cancer with metastases in the lungs.

The aim of this study is to analyze the proteome of the EBCs of patients who were diagnosed with lung cancer in the first/second stage and selection of potential biomarkers for early disease stages.

MATERIALS AND METHODS

Donors. Seventy-nine people were surveyed: 17 patients with chronic obstructive pulmonary disease (COPD) in the acute stage, 13 patients with community-acquired pneumonia, 26 patients diagnosed with lung cancer, and 23 healthy non-smoking volunteers. Patients with COPD and pneumonia were hospitalized in the Pulmonology Department of Moscow City Clinical Hospital No. 57. Diagnosis of COPD and pneumonia was carried out based on generally accepted recommendations, viz., the WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) (http://www.goldcopd.org) and Guidelines for the management of adult lower respiratory tract infections (Joint Taskforce of the European Respiratory Society and European Society for Clinical Microbiology and Infectious Diseases) [12]. The patients who were diagnosed with lung cancer were hospitalized in the Thoracic Surgery Department of the Herzen Moscow Oncology Research Institute; the diagnosis was carried out based on computed tomography of the chest and the results of biopsy analysis. The characteristics of donor groups are shown in Table 1.

This study of exhaled breath condensate was approved by the Ethics Committees of the Emanuel Institute of Biochemical Physics, the Research Institute of Pulmonology, and the Herzen Moscow Oncology Research Institute; all the donors signed informed consent to participate in the study.

Exhaled breath condensate sampling and chromatography-mass spectrometry analysis (HPLC-MS/MS). Exhaled breath condensate of patients who were diagnosed with COPD and community-acquired pneumonia was collected using a stationary ECoScreen device according to previously published protocols [13]. In order to facilitate the procedure for patients who were diagnosed with lung cancer and not disturb them, EBC sampling was performed using a portable RTube device according to previously published protocols [14]. Comparative analysis of the protein lists that were obtained using the ECoScreen and RTube devices was carried out for the control group. Finally, it was shown that the type of device does not influence the final result (data not shown). Sample preparation for mass spectrometry and chromatography-mass spectrometry analysis (HPLC-MS/MS) was performed according to previously published protocols [13, 14].

Bioinformatics data analysis. The list of the exact mass values for the peptides and their fragments was used to search for and identify the proteins in the database using Mascot software (Matrix Science, London, UK; version 2.0.04). For the identification of proteins the IPI-human database (version 3.82; released April 6, 2011; 92 104 entries), which was provided by the European Bioinformatics Institute, was used with the following search parameters: enzyme, trypsin; mass accuracy for the parent ion, 5 ppm; mass accuracy for MS/MS fragments, 0.50 Da; and modification, oxidation of methionine. It was taken that the protein was reliably identified if at least two unique peptides (Score > 70) were found for it in one of the donors, as well as in the case where the protein was found in several donors of the considered group in the presence of at least one unique peptide (Score > 30).

The GeneCards (http://www.genecards.org), GeneOntology (GO) (http://geneontology.org), MOPED (https://www.proteinspire.org/MOPED), BioGPS (http://biogps.org), and UniProt (www.uniprot. org) bioinformatics databases were used for the annotation and analysis of the results; as well, Qiagen products (http://www.qiagen.com) were used for profiling a variety of physiological and pathological processes in the human body.

Table 2. Proteins identified in EBC proteomes of all the donor groups, including the control group

Function	Protein description									
	gene	protein	Mr[1], Da	family[a]	cellular localization[b]		expression in tissues[c/d]			possible association with clinical signs[f]
					external	interna	blood	skin	L&RT[e]	
Enzymes	PTGDS	Prostaglandin-H2 D-isomerase	21029	Lipocalin	+	+	+/+	+/+	–/+	–
	TXN	Thioredoxin	11737	Thioredoxin	+	+	+/+	+/+	+/+	–
Regulatory proteins	AMBP	Alpha-1 -Microglobulin/Bikunin Precursor	38999	Lipocalin/–	+	+	+/+	+/+	–/+	–
	CSTA	Cystatin-A	11006	Cystatin	–	+	+/+	+/+	+/+	–
	DMKN	Dermokine	47082	Dermokine	+	–	+/+	+/+	–/+	–
	KNG1	Kininogen-1	71957	–	+	+	+/+	+/+	–/+	–
Transport proteins	AGP2 (ORM2)	Alpha-1-Acid Glycoprotein 2	23603	Lipocalin	+	–	+/+	+/+	–/+	–
	ALB	Serum albumin	69367	ALB/AFP/VDB	+	+	+/+	+/+	+/+	–
	LCN1	Lipocalin-1	19250	Lipocalin	+	–	+/+	+/+	+/+	–
Structural proteins	DSP	Desmoplakin	331774	Plakin	–	+	+/+	+/+	+/+	–
	FLG2	Filaggrin-2	248073	SlOO-fused protein	–	+	+/+	+/+	+/+	–
	HRNR	Hornerin	282390	SlOO-fused protein	+	+	+/+	+/+	+/+	–
	JUP	Junction plakoglobin	81745	Beta-catenin	–	+	+/+	+/+	+/+	–
	SHROOM3	Protein Shroom3	216857	Shroom	–	+	+/–	+/–	–/–	–
Protective proteins	AZGP1	Zinc-alpha-2-glycoprotein	34259	MHC class I	+	–	+/+	+/+	+/+	–
	DCD	Dermcidin	11284	–	+	–	+/+	+/+	+/+	–
	IGHA1	Immunoglobulin Heavy Constant Alpha 1	37655	–	+	–	+/+	+/+	+/+	–
	LYZ	Lysozyme C	16537	Glycosyl hydrolase 22	+	+	+/+	+/+	+/+	–
	SPP1	Osteopontin	35423	Osteopontin	+	–	+/+	–/+	–/+	–
Contractile proteins	ACTG1	Actin, cytoplasmic 2	41793	Actin	+	+	+/+	+/+	+/+	–
	DNAH14	Dynein heavy chain 14, axonemal	399895	Dynein heavy chain	–	+	+/+	+/+	–/+	–

[a] Molecular weight according to the UniProt database; [b] according to the GeneCards database; [c/d] according to the MOPED/BioGPS database; [e] lungs and respiratory tract; [f] biological/pathological processes according to GO/Qiagen™.

Table 3. Proteins identified only in EBC proteomics of the donors diagnosed with the first/second stage of lung cancer and proposed as a diagnostic panel

Function	Protein description									
	gene	Protein	Mr[1], Da	family[a]	cellular localization[b]		expression in tissues[c/d]			possible association with clinical signs[f]
					external	interna	blood	skin	L & RT[e]	
1	2	3	4	5	6	7	8	9	10	11
Enzymes	PANK2	Pantothenate kinase 2, mitochondrial	62681	Type II pantothenate kinase	–	+	+/+	+/+	–/+	+
	DPYSL2	Dihydropyrimidinase-related protein 2	62294	DHOase	+	+	+/+	+/+	+/+	+
	DPYSL5	Dihydropyrimidinase-related protein 5	61421	DHOase	–	+	+/+	–/+	+/+	+
	DDX20	Probable ATP-dependent RNA helicase DDX20	92241	DEAD box helicase	–	+	+/+	–/+	+/+	+
	SEPT7	Septin-7	50680	Septin GTPase	+	+	+/+	+/+	+/+	+
Regulatory proteins	ATPIF1	ATPase inhibitor, mitochondrial	12249	ATPase inhibitor	–	+	+/+	+/+	–/+	+
	TBC1D1	TBC1 domain family member 1	133084	–	–	+	+/+	+/+	–/+	+
	TBC1D4	TBC1 domain family member 4	146563	–	–	+	+/+	+/+	–/+	+
	NUCKS1	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	27296	–	–	+	+/+	+/+	+/+	+
	POTEE	POTE ankyrin domain family member E	121363	In the N-terminal section belongs to the POTE family; in the C-terminal section belongs to the actin family	+	+	+/na	+/na	+/na	+
	SFRS1	Serine/arginine-rich splicing factor 1	27745	Splicing factor SR	+	+	+/+	+/+	+/+	+
	SFRS3	Serine/arginine-rich splicing factor 3	19330	Splicing factor SR	–	+	+/+	+/+	+/+	+
	SFRS4	Serine/arginine-rich splicing factor 4	56678	Splicing factor SR	–	+	+/+	–/+	+/+	+
	SFRS6	Serine/arginine-rich splicing factor 6	39587	Splicing factor SR	–	+	+/+	–/+	+/+	+
	WDR13	WD repeat-containing protein 13	53696	–	–	+	+/+	–/+	–/+	+
Transport proteins	SYN1	Synapsin-1	74111	Synapsin	–	+	–/+	–/+	–/+	+
Structural proteins	SPDL1	Protein Spindly	70172	Spindly	–	+	+/+	–/+	–/+	+
	BSDC1	BSD domain-containing protein 1	47163	–	–	+	+/+	–/+	–/+	+
	HSP90AA1	Heat shock protein HSP 90-alpha	84660	Heat shock protein 90	+	+	+/+	+/+	+/+	+

[a] Molecular weight according to the UniProt database; [b] according to the GeneCards database; [c/d] according to the MOPED/BioGPS database; [e] lungs and respiratory tract; [f] biological/pathological processes according to GO/Qiagen™.

RESULTS AND DISCUSSION

Proteomic analysis of the EBC of the four presented donor groups revealed more than 300 different proteins. Type II cytoskeletal keratins (1, 2, 3, 4, 5, and 6) and type I cytoskeletal keratins (9, 10, 14, 15, and 16) were invariant for all the samples. In the preliminary experiments we and other authors showed that cytoskeletal keratins are the major protein components of the EBC for smokers, as well as for non-smoking healthy individuals [13, 16]. According to the human-protein catalog, cytokeratins CK 1, 2, 9, and 10 have an epidermal origin; thus, it is believed that these proteins are obtained during sample preparation or have an exogenous origin, i.e., they are not among the respiratory proteins. In a previous study [13, 17], we assumed that these exogenous keratins in EBC are freely circulating proteins in the air.

The complete list of non-keratin proteins that were identified in the samples of the control group and the groups with COPD, community-acquired pneumonia, and lung cancer is given in Table 2.

It can be seen from Table 2 that 21 proteins turned out to be common to all the considered donor groups. The highest frequency of occurrence in the samples was found for dermicidin, which is an antibiotic peptide with antimicrobial and proteolytic activities. It is known that dermicidin is secreted by sweat glands in humans [18, 19], while a number of studies have shown that dermicidin and its derivative peptides are present under acute coronary heart disease in blood [20], placenta [21], brain, and neuronal cell lines [21]. According to the MOPED database, dermicidin expression is observed in many tissues, including the lungs and respiratory tract. Moreover, dermicidin was suggested as a possible oncogene in studies on oncological diseases [22]. It was also shown that it stimulates proliferation of tumor cells in mice, rats and humans [22–24].

The data on the EBC proteomes of the patients with the first and second stages of lung cancer were used for further analysis, because they are of the greatest interest from the diagnostic and prognostic points of view [9].

In the EBC of the donors who were diagnosed with lung cancer in the first/second stage, 42 proteins of a non-keratin origin, which are absent in the EBC of healthy non-smoking control group, as well as in the EBC of the donors with COPD and community-acquired pneumonia, were identified.

Based on detailed annotation using bioinformatics resources, as well as the analysis of the frequency of occurrence, 19 proteins that may be proposed as a diagnostic panel for lung cancer via study of exhaled breath condensate were highlighted (Table 3). As seen from the data that are given in Table 3, the majority of the represented proteins are classified as regulatory proteins according their functional characteristics and have intracellular localization. Expression of all of the represented proteins is increased in this disease; some of them are present in commercial PCR tests on lung cancer profiling in blood (Qiagen™).

It has been shown [25] that the presence of the POTE ankyrin domain family member E is typical for many types of cancer (while it is practically absent in normal tissues). Subsequent studies on the expression of this protein, including in lung cancer [26], have confirmed its promise as a tumor marker; POTE ankyrin domain family member E has been proposed as a target for the development of a cancer vaccine [27]. Serine/arginine-rich splicing factor 1 also often attracts the attention of researchers as a marker of the tumor process [28, 29].

It should be noted that among the proteins that have been identified in the EBC of cancer patients there are many proteins that are involved in mitosis, as well as transcription, translation, and alternative splicing, which may reflect a process of uncontrolled division of tumor cells (BSD domain-containing protein, Heat shock protein HSP 90-alpha, Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1, Probable ATP-dependent RNA helicase DDX20, Protein Spindly, Septin-7, Serine/arginine-rich splicing factor 1–6, WD repeat-containing protein 1). Identification of the group of splicing factors (SR family), which play a critical role in tumor development, is of a great interest [30].

In addition to the represented panel of 19 proteins, the presence of the HMG-I/Y protein family and lactoferrin in the EBC of donors with lung cancer can be noted; this may be associated with capillary network overgrowth that are induced by tumors and the immune response to the activity of cancer cells [14].

Based on the analysis results of the exhaled breath condensate proteome it can be stated that the protein profiles of the different donor groups can be distinguished and there is a chance to highlight a specific protein group typical for certain condition/disease of the respiratory system. It should be noted that the proteome of cancer patients that was determined in the EBC is very different from the proteomes of not only healthy young non-smoking control group, but also from the proteomes of the patients with COPD and pneumonia in the same age group, indicating the potential use of the EBC as a screening test followed by verification using other methods, such as computed tomography.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project no. 15-04-05168 A). The high-resolution mass spectrometry study was financially supported by the Russian Foundation for Basic Research and the Government of Moscow in the framework of research project no. 15-38-70039 mol_a_mos and international project

no. 15-58-52041 NNC_a of the Russian Foundation for Basic Research.

REFERENCES

1. Horváth, I., et al., *Eur. Respir. J.*, 2005, vol. 26, no. 3, p. 523.
2. Buszewski, B., et al., *Biomed. Chromatogr.*, 2007, vol. 21, no. 6, p. 553.
3. Kurova, V.S., et al., *Russ. Chem. Bull.,* 2010, vol. 59, no. 1, p. 292.
4. Horváth, I., et al., *Eur. Respir. J.*, 2009, vol. 34, no. 1, p. 261.
5. Szymanski, W.W., et al., *Meas. Sci. Technol.*, 2002, vol. 13, no. 3, p. 303.
6. Ferlay, J., et al., *Int. J. Cancer*, 2010, vol. 127, no. 12, p. 2893.
7. Jemal, A., et al., *CA: Cancer J. Clin.*, 2008, vol. 58, no. 2, p. 71.
8. Agrawal, A., et al., *J. Carcinog.*, 2013, vol. 12, no. 1, p. 3.
9. Reed, M.F., et al., *Am. J. Surg.*, 2004, vol. 188, no. 5, p. 598.
10. Cuzick, J., *Eur. J. Cancer*, 1999, vol. 35, no. 5, p. 685.
11. Brothers, J.F., et al., *BMC Med.*, 2013, vol. 11, no. 1, p. 168.
12. Woodhead, M., et al., *Clin. Microbiol. Infect.*, 2011, vol. 17.
13. Kurova, V.S., et al., *Clin. Chem. Lab. Med.*, 2009, vol. 47, no. 6, p. 706.
14. Ryabokon', A.M., et al, *Pul'monologiya*, 2014, no. 1, p. 5.
15. Cheng, Z., et al., *J. Cancer Ther.*, 2011, vol. 02, no. 01, p. 1.
16. Hoffmann, H.J., et al., *Eur. Respir. J.*, 2008, vol. 31, no. 2, p. 380.
17. Kurova, V.S., et al., *Russ. J. Bioorg. Chem.*, 2011, vol. 37, no. 1, p. 48.
18. Ghosh, R., Jana, P., and Sinha, A., *Exp. Clin. Endocrinol. Diabetes*, 2012, vol. 120, no. 03, p. 145.
19. Ghosh, R., et al., *Thrombosis*, 2012, p. 987932.
20. Mikhaylova, M., et al., *BMC Res. Notes*, 2014, vol. 7, no. 1, p. 400.
21. Cunningham, T.J., et al., *J. Neurosci.*, 1998, vol. 18, no. 18, p. 7047.
22. Stewart, G.D., et al., *Br. J. Cancer*, 2008, vol. 99, no. 1, p. 126.
23. Todorov, P., et al., *Nature*, 1996, vol. 379, no. 6567, p. 739.
24. Park, S.-Y., et al., *Arch. Pharm. Res.*, 2010, vol. 33, no. 2, p. 247.
25. Bera, T.K., et al., *Cancer Res.*, 2006, vol. 66, no. 1, p. 52.
26. Wang, Q., et al., *PLoS One*, 2015, vol. 10, no. 4.
27. Huang, Y.-H., et al., *PLoS One*, 2013, vol. 8, no. 6.
28. Ezponda, T., et al., *Clin. Cancer Res.*, 2010, vol. 16, no. 16, p. 4113.
29. De Miguel, F.J., et al., *Cancer Res.*, 2014, vol. 74, no. 4, p. 1105.
30. Da Silva, M.R., et al., *Biomed. Res. Int.,* 2015, vol. 2015, p. 150514.

Translated by D. Novikova