METABOLOMIC STUDY OF SERUM, URINE, AND BRONCHOALVEOLAR LAVAGE FLUID BASED ON GAS CHROMATOGRAPHY-MASS SPECTROMETRY TO DELVE INTO THE PATHOLOGY OF LUNG CANCER

Belén Callejón-Leblic[a,b,c], Tamara García-Barrera[*a,b,c], Antonio Pereira-Vega[d], José Luis Gómez-Ariza[*a,b,c]

[a]Department of Chemistry, Faculty of Experimental Sciences. University of Huelva. Campus de El Carmen. 21007, Huelva. SPAIN; [b]International Agrofood Campus of Excellence International ceiA3. University of Huelva. SPAIN; [c]Research Center of Natural Resources, Health and Environment (RENSMA). University of Huelva. Campus de El Carmen. 21007, Huelva. SPAIN; [d]Pneumonology Area of Juan Ramón Jiménez Hospital, Huelva, Spain

Corresponding authors at:
Department of Chemistry. Faculty of Experimental Science. University of Huelva. Campus de El Carmen. 21007, Huelva. Tel.: +34959219968, fax: +34 959 219942, e-mail address: ariza@uhu.es (J.L. Gómez-Ariza)

Department of Chemistry. Faculty of Experimental Science. University of Huelva. Campus de El Carmen. 21007, Huelva. Tel.: +34959219962, fax: +34 959 219942, e-mail address: tamara@dqcm.uhu.es (T. García-Barrera)

Abstract

This study explores for the first time the combination of serum, urine and bronchoalveolar lavage fluid (BALF) to deep insight into the pathology of lung cancer (LC) using a metabolomic platform based on gas chromatography mass spectrometry (GC-MS). The study includes LC

patients, a healthy control group (HC) and a group of patients with noncancerous lung diseases (NCC) used as a control group respect to BALF because of the invasive nature this fluid collection.

The metabolomic platform was applied to serum, urine and BALF samples in order to compare the metabolomic profiles of these biological fluids and establish metabolic similarities and differences between them. The application of PLS-DA presented a clear classification of groups for all types of samples, indicating the existence of altered metabolites in LC. Twenty-six and thirty-one perturbed metabolites in the LC were annotated in the comparison of serum and urine samples. On the other hand, sixteen metabolites were altered in BALF of LC patients compared to NCC. The pathway analysis indicated that several amino acid metabolic routes were the most affected in LC. Finally, ROC curves were applied to the dataset and metabolites with an AUC value higher than 0.75 were considered as relevant in the progression of LC.

Keywords: Serum, Urine, Bronchoalveolar lavage, Lung cancer, Gas chromatography, Metabolomics.

1. Introduction

It is well-known that the aggressive character of lung cancer (LC) causes more deaths than the combination of breast, prostate, colon and pancreatic cancers together [1]. LC is an unspecific disease difficult to discriminate from other lung disorders that leads to late diagnosis, being necessary to have safe and early diagnosis methods that allow to increase the averaged survival period of patients longer than five-years from 15% of cases to 85 % [2]. Several biomarkers have been extensively used for the diagnosis of LC as carcinoembryonic antigen (CEA), cancer antigen 125, cytokeratin 19 fragment (CYFRA21-1), and neuron-specific enolase [3], but these biomarkers present low accuracy, which limit their utility for early detection of LC.

Screening techniques, such as low-dose computed tomography (LDCT) could reduce mortality due to LC by 20%, however, the incidence of high false-positive rates (96%) decrease the efficacy of the technique and complementary noninvasive and sensitive/specific biomarkers should be used in conjunction with LDCT for more accurate diagnosis [4]. Other radiologic diagnosis techniques, such as ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography FDG (PET/CT), have proved a high capability to detect cancerous lesions in the lung, despite their high cost makes these tests unable for general public annual screening under any healthcare system.

Consequently, a deeper knowledge of metabolites involved in the normal physiological function of human cells and organs and the suitable interpretation of theirs interactions is essential to understand pathological mechanisms of diseases and the searching of early diagnostic markers, and metabolomics is a very suitable tool for this purpose [5]. Metabolomics has emerged as a valuable complement to genomics and proteomics to provide information related to cellular metabolic processes that drive tumor formation and progression. Metabolomics allows evaluating the variability in the number and type of metabolites expressed in lung tumors, which are related to the type, stage, and, potentially, the response to drug treatment [6].

Therefore, the use of metabolic biomarkers to supplement imaging diagnosis is an interesting alternative to detect the onset of the disease, and the application of non-targeted metabolomic techniques for this purpose can provide new options in this field, since it can monitor the changes of these key-biomolecules in response to LC.

Several metabolomic platforms has been proposed in searching LC biomarkers based on NMR [7] and MS [2]. Generally, MS metabolomic approaches use couplings with HPLC [7] capillary electrophoresis [8] and gas chromatography [2,9], although direct infusion of the sample into the mass spectrometer has also been proposed [10].

Several human biofluids and tissue biopsies have been used in metabolomic studies[5] aimed to LC diagnosis, such as serum [2,11,12], plasma [13,14], urine [15], sputum [16], and lung tissues [2,17], but only a few authors had described the combined use of fluids or tissues for LC diagnosis purpose [2]. This complementary approach is useful because similar changes in both biofluids and resected tumors from cancer patients can reflect metabolic changes in the tumor itself. In addition, during oncogenesis the levels of some metabolites increase releasing into the blood, which results in higher levels of these metabolites in the serum. On the contrary, certain metabolites can be transferred from the blood to lung tumor promoting cancer proliferation [6]. So, the combined study of tumor tissues and biofluids allows not only the search for biomarkers for early diagnosis, but also as indicators of disease progression.

In connection to this, bronchoalveolar lavage is a fluid obtained during the exploratory study of lung patients (bronchoscopy) through instillation and later aspiration of liquid in one or more lung segments, which provide information about cellular and epithelial surface of the lower respiratory tract. It is estimated that BALF samples take a million cells (1% of the lung surface), providing about 1 ml of pulmonary secretions in the total recovered liquid [18]. Since BALF is in close interaction with lung tissue is a very representative sample of lung status, and presents some equivalence with lung biopsies.

Previous metabolomic studies based on BALF samples consider different lung diseases, such as asthma [19] or cystic fibrosis [20], but only one study reported by Callejón-Leblic et al. [21] focus on the use of this fluid in LC diagnosis.

The aim of the present study was to identify altered metabolites in three different fluids from LC patients, which could provide interesting information about the pathology of LC and the proposal of tentative biomarkers for early diagnosis of the disease. To our knowledge, this is the first study which presents the joint study of these three fluids (serum, urine and BALF) in the diagnosis of LC using a metabolomic platform based on gas chromatography mass spectrometry (GC-MS).

2. *Material and methods*

2.1. Sample collection

Blood and urine samples were collected from 32 lung cancer patients and 29 healthy people (HC) at the Pneumonology Area of Juan Ramón Jiménez Hospital (Huelva, Spain) from July 2011 to July 2013. In the same way, a total of 54 BALF samples (24 LC and 30 NCC) obtained by bronchoscopy, were collected. Because the invasive character of the technique used to obtain the BALF, it is not possible to get samples from healthy people. In addition, the 24 BALF samples from LC patients overlapped to 24 serum and urine samples from the same patients. The blood samples were obtained by venipuncture of the antecubital region, after 8 hours of fasting, and collected in BD Vacutainer SST II tubes with gel separator and Advance vacuum system. The samples were immediately cooled and protected from light for 30 minutes to allow clot retraction. After centrifugation (2057 g for 10 minutes) serum, urine and BALF samples were aliquoted in Eppendorf tubes and frozen at -80°C until analysis. Clinical data of LC, NCC and HC patients are shown in Supplementary Material (Table 1 from serum and urine samples and in Table 2 from BALF samples).

The work was performed in accordance with the principles contained in the Declaration of Helsinki and approved by the Ethical Committee of Juan Ramón Jiménez Hospital and University of Huelva.

2.2. Reagents

All the solvents used were of HPLC-grade. Methanol, ethanol and pyridine were purchased from Aldrich (Steinheim, Germany). Formic acid was supplied by Merck (Darmstadt, Germany) and derivatizing agents, namely methoxylamine hydrochloride and N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA), as well as urease from *Canavalia ensiformis* (Jack bean) Type IX powder 50000-100000 units/g solid were obtained from Aldrich (Steinheim, Germany). Water

was purified with a Milli-Q Gradient system (Millipore, Watford, UK).

2.3. Sample treatments

Serum samples were extracted following a procedure described by Bruce et al. [22]. In brief, 100 µL of serum were mixed with 400 µL of 1:1 MeOH/EtOH mixture in an Eppendorf tube and vortexed for 5 min at room temperature followed by centrifugation at 2057 g for 10 min at 4°C to discard the pellet which contains the protein fraction. The supernatant was transferred to another Eppendorf tube and was dried with a speed vacuum system (Thermo Scientific Savant® SPD111V SpeedVac® Concentrator) at 30 °C for 20 min.

For urine sample treatment, 10 µl of urease from *Canavalia ensiformis* (Jack bean) Type IX (1 mg/µl) were added to 50 µl of centrifuged urine and the mixture was incubated at 37°C during 10 min. Then, 200 µl of MeOH was mixed with the sample and vortexed for 5 min. Urines were dried with a speed vacuum system in the same condition of serum samples.

Finally, 500 µl of BALF samples were pre-concentrated using a speed vacuum system during 1h at 45 ºC. Serum, urine and BALF samples were reconstituted with the derivatizing agents. For the protection of carbonyl groups by methoxymation, dried extracts were redissolved in 50 µl of 20 mg mL^{-1} methoxyamine in pyridine, and after briefly vortexing were incubated at 80◦C for 15 min in a water bath. Subsequently, silylation was performed by adding 50 µl of MSTFA and incubation at 80◦C for 15 min. Finally, extracts were centrifuged at 2057 g for 5 min and the supernatant collected for analysis.

2.4. Quality controls

For each matrix, a total of seven Quality Control samples (QCs) were prepared by pooling equal volumes of all samples studied, which were treated with the sample procedure and analyzed for GC-MS. In brief, 20 µl of serum from all patients of the study (32 LC and 29) were mixed to prepare the QCs samples used in the analysis of serum. In the same way, 20 µl of urine from all

patients (32 LC and 29 HC) and BALF (24 LC and 30 NCC) were pooled independently for the metabolomics analysis of urine and BALF, respectively. Quality controls results were statistically treated by principal component analysis (PCA) and score plots represented in order to check the stability during the analysis.

2.5. GC-MS analysis.

Gas chromatographic analysis was performed in a Trace GC ULTRA gas chromatograph coupled to an ion trap mass spectrometer detector ITQ900 (Thermo Fisher Scientific), using a Factor Four capillary column VF-5MS 30 m × 0.25 mm ID, with 0.25 µm of film thickness (Varian). The GC column temperature was set to 100ºC for 0.5 min, and programmed to reach 320ºC at a rate of 15ºC per minute. Finally, this temperature was maintained for other 7 min, being the total time of analysis 22.17 min. The injector temperature was kept at 280ºC, and helium was used as carrier gas at a constant flow rate of 1 ml min^{-1}. For mass spectrometry detection, ionization was carried out by electronic impact (EI) with a voltage of 70 eV, by full scan mode in the m/z range 35–650, with anion source temperature of 200ºC. For analysis, 1 µl of sample was injected in splitless mode.

2.6. Data processing

Raw data were processed following the pipeline described by Katajamaa and Oresic [23], which proceeds through multiple stages including feature detection, alignment of peaks and normalization. For this purpose the freely available XCMS software, included in the R platform (http://www.r-project.org) was used. Files were converted into net CDF using the Thermo File Converter tool (Thermo Fisher Scientific) and subsequently, data were extracted using the matched filter method. This algorithm slices data into extracted ion chromatograms (XIC) on a fixed step size, and then each slice is filtered with matched filtration using a second-derivative Gaussian as the model peak shape. The XCMS parameters were optimized according to the

characteristics of datasets obtained in order to extract the maximum information as possible. Finally, the settings applied for GC-MS data were S/N threshold 2, full width at half-maximum (fwhm) 3, and width of the m/z range 0.1 (step parameter). After peak extraction, grouping and retention time correction of peaks (alignment) was accomplished in three iterative cycles with descending bandwidth (bw) from 5 to 1 s. For data normalization, the locally weighted scatter plot smoothing (LOESS) normalization method was used, which adjusts the local median of log fold changes of peak intensities between samples in the data set to be approximately zero across the whole peak intensity range. The preprocessed data were then exported as a .csv file for further data analysis by multivariate procedures.

2.7. Statistical Analysis and biomarker identification

GC-MS data were processed with SIMCA-P™ software (version 11.5, published by UMetrics AB, Umeå, Sweden) to perform principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) in order to compare metabolomic profiles obtained. Before performing statistical analysis, data was submitted to Pareto scaling, for reducing the relative importance of larger values, and logarithmic transformation, in order to approximate to a normal distribution. Finally, potential biomarkers were selected according to the Variable Importance in the Projection, or VIP (a weighted sum of squares of the PLS weight, which indicates the importance of the variable in the model), considering only variables with VIP values higher than 1, indicative of significant differences among groups. NIST Mass Spectral Library (version 08) was used to identify the altered metabolites, considering only those variables with a probability greater than 80%. On the other hand statistical significance was analyzed using one-way ANOVA followed by multiple tests correction with the Tukey test (STATISTICA 8.0 sotfware, StatSoft, Tulsa, USA), using a level of probability of 0.05 as criterion for significance.

2.8. Metabolic pathway analysis

Metabolic pathway analysis was performed to identify the affected pathways on the basis of altered metabolites. The Metaboanalyst 4.0 web tool (http://www.metaboanalyst.ca/) was used for this purpose, which conducts pathway analysis through pathway enrichment analysis and pathway topological analysis. The "Homo sapiens" library was selected using the default "Hypergeometric Test" and "Relative-Betweenness Centrality" algorithms for pathway enrichment analysis and pathway topological analysis, respectively. In order to identify the most relevant pathways, the impact-value threshold calculated from pathway topology analysis was set at 0.1.

2.9. Sensitivity and Specificity of altered metabolites in serum, urine and BALF

The area under the curve of the receiver operator characteristic (ROC) was used to determine the specificity and the sensitivity of altered metabolites in relation to the disease to discriminate patients with lung cancer from control subjects. The analysis was applied to the resulting altered metabolites in LC detected from GC-MS using the Metaboanalyst 4.0 (http://www.metaboanalyst.ca/) software. Resulting data was carried out by logarithmic transformation and Pareto scaling.

3. Results

The resulting mass data of serum and urine samples from LC and HC and mass data of BALF samples from LC and NCC were subjected to multivariate statistical analysis. Principal component analysis (PCA) was applied for a preliminary evaluation of the quality of the data before the application of any supervised statistical analysis. In addition, grouping of quality control from serum, urine and BALF samples in PCA corroborated the stability of the GC-MS analysis (Supplementary Material, Figure 1). The PLS-DA score plots from serum (Figure 1a) and urine samples (Figure 1b) showed a good classification between LC and HC groups and

234 presented good validation parameters of the model, being the statistic performance R^2Y=0.986
235 and the predictive parameter Q^2= 0.731 for serum and R^2Y=0.982 and Q^2= 0.777 for urine. A
236 total of 26 altered metabolites (Table 1) were annotated in serum from LC patients. In
237 addition, p-values from one-way ANOVA followed by a multiple test correction based on Tukey
238 test was carried out in the serum and urine dataset. AUC values of ROC curves analysis to
239 determine the specificity and sensitivity of altered metabolites in LC are also shown in Table 3.
240 L-valine, L-glycine, uridine, malonic acid, fructose, phosphoric acid, L-asparagine, deoxy-
241 glucose, glucose uric acid, stearic acid and margaric acid showed AUC values higher than 0.75
242 suggesting being possible biomarkers of LC.
243
244 Table 1. Altered metabolites in serum from LC with VIP, sense of alteration, fold change (LC/HC), p-value
245 from one-way ANOVA followed by Tukey test and AUC values.
246

Metabolite	Rt (min)	VIP	Disturbance sense	Fold change	*p* values	AUC
L-Valine*	4.15	1.5	↓	0.53	0.001	0.75
L-Glycine*	4.63	1.33	↓	0.91	0.022	0.76
Tartaric acid	5.02	1.2	↓	0.54	0.01	0.72
L-Serine	5.32	1.12	↓	0.67	0.012	0.67
L-Threonine	5.48	1.38	↑	1.54	0.009	0.69
Uridine*	5.82	1.31	↓	0.6	0.004	0.76
Malonic acid*	6.23	1.02	↑	1.02	0.002	0.75
L-Proline	6.45	1.39	↓	0.64	0.002	0.71
L-Cysteine	6.92	1.28	↑	1.9	0.039	0.64
L-Glutamine	7.40	1.57	↓	0.61	0.043	0.67
L-Phenilalanine	7.53	1.92	↑	1.03	0.022	0.73
Fructose*	7.57	2.24	↓	0.69	0.009	0.75
Phosphoric acid*	8.35	1.63	↓	0.82	0.005	0.77
Isocitric acid	8.57	2.27	↓	0.77	0.013	0.61
L-Asparagine*	8.77	2.57	↓	0.52	0	0.83
Inositol	8.87	1.66	↓	0.56	0.002	0.74
L-Ornithine	8.97	1.32	↓	0.7	0.018	0.65
Deoxy-glucose*	9.10	2.31	↑	4.24	0.046	0.8
Glucose*	9.38	2.74	↑	2.29	0.001	0.82
Palmitic acid	10.18	2.01	↓	0.94	0.037	0.53
Uric acid*	10.72	2.64	↓	0.38	0.003	0.79
Stearic acid*	11.38	1.82	↓	0.41	0.001	0.8
L-Cystine	11.88	1.48	↓	0.91	0.009	0.54
Myristic acid	12.45	2.42	↓	0.5	0.042	0.73

Margaric acid*	13.93	2.24	↑	0.42	0.012	0.78
Arachidonic acid	15.30	1.98	↓	0.65	0.02	0.67

*Metabolites with AUC higher than 0.75

Sample treatment for urine metabolome analysis was optimized to reduce urea interference in mass spectra by adding urease from *Canavalia ensiformis* (Jack bean) Type IX. Figure 2 compares two chromatograms obtained from the analysis of urine samples with and without sample treatment with urease. It can be observed that the band generated between 2 and 3 minutes disappears with the use of the enzyme. Table 2 showed a total of 32 altered metabolites in LC resulting of the comparison between LC and HC. The values of VIP, Fold change, disturbance sense, p-values from one-way ANOVA followed by a multiple test correction based on Tukey test and AUC values from ROC curves analysis are also shown in Table 2. In urine, the perturbed metabolites with AUC values higher than 0.75 were malonic acid, L-glycine, succinic acid, L-threonine, adipic acid, aconitic acid, phosphoric acid, isocitric acid, inositol, glucose, glucaric acid, uric acid and stearic acid.

Table 2. Altered metabolites in urine from LC with VIP, sense of alteration, fold change (LC/HC), p-value from one-way ANOVA followed by a multiple test correction Tukey test and AUC values.

Metabolite	Rt (min)	VIP	Disturbance sense	Fold change	*p* values	AUC
L-Alanine	3.30	1.19	↑	8.6	0.031	0.59
Acetic acid	3.38	1.72	↑	3.17	0.043	0.59
Malonic acid*	3.67	1.73	↑	3.31	0.005	0.79
Urea	3.82	1.85	↑	8.84	0.011	0.7
L-Glycine*	4.63	2.46	↓	0.29	0.005	0.82
Succinic acid*	4.88	1.42	↑	7.21	0.033	0.81
Glyceric acid	4.95	1.79	↑	4.12	0.013	0.74
L-Serine	5.32	1.29	↑	2.71	0.033	0.61
L-Threonine*	5.48	1.74	↑	4.07	0.029	0.82
Butanoic acid	5.78	1.52	↑	5.17	0.01	0.68
Threonic acid	6.67	1.93	↑	2.03	0.009	0.7
Creatinine	6.90	1.94	↑	1.71	0.029	0.67
Glutaconic acid	7.10	1.28	↑	1.84	0.003	0.73
L-Aspartate	7.48	1.23	↑	2.41	0.047	0.73
Ribonic acid	7.57	2	↑	2.91	0.043	0.66
Adipic acid*	7.78	2.21	↑	3.96	0.017	0.75

Arabitol	7.90	1.36	↑	2.72	0.02	0.63
Ribitol*	8.08	1.29	↑	2.76	0.003	0.79
Aconitic acid*	8.27	1.51	↑	2.09	0	0.81
Phosphoric acid*	8.35	1.61	↓	0.39	0.011	0.75
Isocitric acid*	8.57	1.44	↓	0.4	0.034	0.75
Hippuric acid	8.78	2.37	↓	0.47	0.035	0.69
Purine	8.82	1.87	↑	3.72	0	0.54
Inositol*	8.87	1.59	↓	0.17	0.005	0.76
Gluconic acid	9.23	2.01	↑	3.35	0.021	0.74
Sorbitol	9.48	1.24	↑	2.37	0.003	0.67
Glucaric acid*	9.55	2.07	↓	0.11	0.048	0.93
Galactaric acid	9.65	1.56	↑	3.26	0.019	0.67
Palmitic acid	10.18	1.53	↑	1.85	0.032	0.7
Uric acid*	10.72	1.99	↑	4.25	0.044	0.75
Stearic acid*	11.38	2.1	↑	2.79	0	0.88

* Metabolites with AUC higher than 0.75

Finally, PLS-DA score plots from BALF samples (LC *vs* NCC) are shown in Figure 1c, exhibiting a statistic performance of R^2Y=0.992 and the predictive parameter Q^2= 0.763. A total of 16 altered metabolites (Table 3) were annotated in LC. The *p* value from one way ANOVA followed by Tukey test and AUC are also collected in Table 3, in this case only glycerol and phosphoric acid showed AUC > 0.75

Table 3. Altered metabolites in BALF from LC with VIP, disturbance sense, fold change (LC/NCC), p-value from one-way ANOVA followed by a multiple test correction Tukey test and AUC values.

Metabolite	Rt (min)	VIP	Sense of alteration	Fold change	*p* values	AUC
Lactic acid	2.87	1.96	↓	0.62	0.012	0.65
Acetic acid	3.38	1.93	↑	1.71	0.047	0.52
Glycerol*	4.28	2.18	↓	0.59	0.004	0.88
L-Glycine	4.63	1.49	↓	0.6	0.017	0.74
L-Aspartate	6.35	1.35	↓	0.48	0.047	0.74
L-Proline	6.45	1.01	↓	0.68	0.042	0.63
L-Glutamine	7.13	1.44	↓	0.50	0.033	0.66
Fructose	7.57	1.66	↓	0.58	0.017	0.60
Phosphoric acid*	8.35	2.12	↓	0.42	0.047	0.79
Isocitric acid	8.57	1.67	↓	0.53	0.023	0.57
Inositol	8.87	1.17	↓	0.63	0.028	0.58
Galactose	9.08	2.1	↓	0.63	0.043	0.68
Palmitic acid	10.18	2.11	↑	1.37	0.008	0.69
Stearic acid	11.38	1.88	↑	1.36	0.022	0.64
Inosine	13.13	1.02	↓	0.36	0.038	0.63

Oleic acid	13.70	1.45	↓	0.78	0.023	0.54

* Metabolites with AUC higher than 0.75.

Subsequently, among total altered metabolites observed in the different analyzed samples (serum, urine and BALF) only six were common in the three fluids, as can be seen in the Euler diagram (Fig. 3), which shows the number of altered metabolites in LC and the coincidences between fluids. On the other hand, seven metabolites were common in serum and BALF, seven in urine and BALF, and ten in serum and urine. Table 4 shows that four of six common metabolites of the three fluids (L-glycine, phosphoric acid, isocitric acid and inositol) presented the same change sense, independently of the fluid considered which is indicative of the role of these metabolites in relation to the pathology of LC. On the contrary, palmitic and stearic acid showed an increase in urine and BALF, and a decrease in serum. On the other hand, perturbed L-serine and malonic acid presence were in common in serum and urine samples, although the change sense for L-serine decreased in serum and increased in urine. However, malonic acid increased in both fluids. Finally, L-proline presented reduced levels in serum and BALF, and L-aspartate increased in urine and decreased in BALF.

Table 4.- Altered metabolites in most than one fluid (serum, urine and BALF)

Metabolites	SERUM	URINE	BALF
Palmitic acid	↓(FC = 0.94)	↑(FC = 1.85)	↑(FC = 1.37)
Phosphoric acid	↓(FC = 0.82)	↓(FC = 0.39)	↓(FC = 0.42)
Isocitric acid	↓(FC = 0.77)	↓(FC = 0.41)	↓(FC = 0.53)
L-Serine	↓(FC = 0.67)	↑(FC = 2.71)	-
L-Proline	↓(FC = 0.64)	-	↓(FC = 0.68)
Inositol	↓(FC = 0.56)	↓(FC = 0.17)	↓(FC = 0.63)
Stearic acid	↓(FC = 0.41)	↑(FC = 2.79)	↑(FC = 1.36)
Uric acid	↓(FC = 0.38)	↑(FC = 4.25)	-
L-Glycine	↓ (FC = 0.91)	↓(FC = 0.29)	↓(FC = 0.60)
L- Threonine	↑(FC = 1.54)	↑(FC = 4.07)	-
Malonic acid	↑(FC = 1.02)	↑(FC = 3.31)	-
L-Aspartate	-	↑(FC = 2.41)	↓(FC = 0.48)

FC: Fold Change

The altered metabolic pathways associated with the metabolites perturbed in LC were identified using the pathway analysis tool from “Metaboanalyst 4.0”. A total of 22, 34 and 40

metabolic routes have been altered in LC from serum, urine and BALF, respectively. A representation of the impact of these routes in LC can be observed in Figure 4 where each dot represents a metabolic pathway and its size indicates the impact of pathways in response to LC. Perturbations in glycine, serine and threonine metabolism, arginine and proline metabolism, inositol and phosphate metabolism, alanine, aspartate and glutamate metabolism, pyruvate metabolism, galactose metabolism and cysteine and methionine metabolism have been observed, being the most altered pathways in LC calculated according to the corresponding p and Impact values. Figure 5 shows a summary chart of the interconnected metabolic pathways related to altered metabolites in LC.

4. Discussion.

In this study, a metabolomic approach based on GC-MS was applied to serum, urine and BALF samples from LC and control patients. The main novelty of this work is the application of metabolomics to three human biological fluids from patients with LC to identify altered metabolites that can help to decipher the pathology of LC and the proposal of tentative biomarkers to the early diagnosis of the disease. In addition, it is the first time that metabolomic relationships between serum, urine and BALF samples from patients with LC are considered.

Statistical analysis PLS-DA from serum and urine samples allowed identifying twenty-six and thirty-one altered metabolites, respectively, in LC. In relation to BALF, sixteen metabolites showed perturbations in LC compared to patients with non-cancerous lung diseases being palmitic acid, phosphoric acid, isocitric acid, inositol, stearic acid and malonic acid altered metabolites in common with serum and urine, as has been previously commented.

ROC curves analysis of the total altered common metabolites in serum, urine and BALF showed that phosphoric acid was the only metabolite with higher AUC values in the three fluids,

suggesting the importance of this metabolite in the pathogenesis of LC. On the other hand, malonic, palmitic, stearic and uric acids presented AUC values higher than 0.75 in serum and urine samples from LC groups. The purpose of relating altered metabolites in different fluids, as well as the evaluation of the specificity and sensitivity of them, is to obtain a set of biomarkers that facilitate the diagnosis of LC.

A number of amino acids change in the three fluids from LC patients (L-glycine, L-alanine, L-valine, L-proline, L-serine, L-asparagine, L-glutamate, L-aspartate, L-threonine, L-ornithine, L-cysteine and L-glutamine) which is in accordance with previous reports from other authors [2,9,12,14,17]. L-glycine, L-proline, L-valine, L-serine, L-asparagine and L-ornithine diminished in serum, and L-glycine diminished in urine; and L-glycine, L-aspartate, L-proline and L-glutamine decreased in BALF samples from LC patients. On the other hand, L-threonine and L-cysteine increased in serum and L-alanine and L-aspartate increased in BALF samples. L-glycine was the only common altered amino acid that decreased in all the fluids. This metabolite is involved in the serine, threonine and glycine metabolism, one of the most altered metabolic route in the pathway analysis (Fig.4) and consequently in the production of succynil-CoA, important substratum in the TCA cycle. Several studies showed a reduction of L-glycine levels in LC patients in accordance to our results, as Chen et al [9], who reported a decrease of this amino-acid in serum from preoperative LC patients using GC-MS and Wen et al [24] who described low levels of L-glycine in plasma from LC adenocarcinoma patients compared to healthy people. On the other hand, a recent metabolomic study in sputum from LC patients stated also a decrease of this metabolite in NSCLC and SCLC patients [25], and Rocha et al described a decrease of this amino acid in human lung tissue from LC patients [17]. In addition, Zhang et al., found an increase of the enzyme glycine descarboxylase (GLDC) in cells isolated from NSCLC tumors with a concomitant decrease in glycine [26]. GLDC is involved in the degradation of glycine which is associated to the generation of methyl groups that can be used in purine biosynthesis. Therefore, the overexpression of this enzyme could be explained by the

decrease of glycine in LC. In addition, L-glycine showed high AUC values in serum (Table 1), urine (Table 2), and BALF (Table 3) suggesting that this metabolite could be considered as possible biomarker of LC. L-proline decreased in serum and BALF samples from LC patients. This metabolite participates in the arginine and proline metabolic pathway and is related to urea cycle and nitrogen metabolism. According to our results, other authors reported that levels of L-proline were reduced in patients with LC adenocarcinoma [24] and postoperative LC patients [9] compared with healthy people. On the other hand, Klupzynska et al., reported a decrease of L-proline in serum from LC patients using UPLC/Q-TOF-MS [27].

L-aspartate increased in urine, but decreased in BALF samples. This non-essential amino acid is involved in the production of proteins and it participates in numerous metabolic routes, being the alanine, aspartate and glutamate metabolism one of the most relevant routes related to TCA cycle. In addition, L-aspartate produces L-asparagine, an amino acid that presents alterations in the present work generating oxaloacetate, which suggests perturbations in energy metabolism. Hori et al. [2] reported decreased levels of L-aspartate in serum, which increased in lung tissue from LC, and Klupzynska et al. [27] found increased levels of this amino acid in NSCLC patients.

The levels of fatty acids such as palmitic, margaric, arachidonic, oleic and stearic acids were found altered in fluids from LC patients, which can be associated to proliferating tumor cells that use long chain fatty acids for membrane assembly, lipid modifications of proteins and energy production. Only palmitic and stearic acid alterations were common in the three fluids but the behavior was different between them, increasing in BALF and urine and decreasing in serum. Yanjie Li et al,[28] reported a decrease of stearic acid in serum, whereas other author as Hori et al., reported an increase of it in lung tumor tissues [2] and Yingrong et al., found increased levels of stearic acid in postoperative lung cancer serum samples compared to preoperative lung cancer patients [9].

Other organic acids as acetic, lactic, phosphoric, malonic, isocitric, uric, butanoic, ribonic, hexanedioic, hippuric and aconitic acids were altered in LC. However only phosphoric and isocitric acids showed similar changes in BALF, serum and urine decreasing in the three fluids. Isocitric acid is related to the formation of α-ketoglutarate also involved in Krebs cycle, and phosphoric acid participates in oxidative phosphorylation, metabolic pathway related to the production of ATP. Hori et al., [2] described a diminution of phosphoric acid in serum from LC patients, similarly to that reported by Chen et al,[9] which is agreement with our results. It should also be noted that the AUC values of phosphoric acid determined in the three fluids (serum, urine and BALF) were greater than 0.75 (Table 1, 2 and 3) in all the cases, supporting the possible association of this metabolite to the progression of LC.

Glycerol is a three-carbon substance that forms the backbone of fatty acids in fats and, consequently, contributes to phospholipids biosynthesis. Glycerol can be converted to glucose by the liver and provides energy for cellular metabolism. Glycerol decreased in BALF from LC patients due to its rapid conversion to glucose, which increased in LC to sustain cancer growth [29]. On the other hand, fructose, galactose, glucose and 2-deoxy-D-glucose**,** involved in carbohydrates metabolism and other acids implicated in energy metabolism as glucaric and galactaric acid decreased in LC also in connection with energy metabolism.

Decreased levels of inositol were found in serum, urine and BALF samples from LC patients. This metabolite is an isomer of glucose being a fundamental component of the cell membrane, so it is an essential nutrient required by human cells for growth and survival in culture and it is involving in inositol phosphate metabolism. Several authors reported a decrease of inositol in LC patients as Chen et al, in preoperative LC patients compared to healthy controls and also in comparison to postoperative LC patients [9], and Rocha et al described a decrease of this metabolite in human lung tissue from LC people, proposing that inositol are reduced in tumors possibly as consequence of local changes in osmotic regulation [17].

5. CONCLUSIONS

The application for the first time of a metabolomic platform to three human biofluids (serum, urine and BALF) of LC patients based on GC-MS has provided new information about the pathogeny of this disease. Despite the novelty of comparing altered metabolites in three fluids from LC patients, this study presents a limitation due the different controls used (HC and NCC). However, the use of this type of control for BALF is the only way to have a metabolomic comparison between people with lung cancer and others who do not have it.

Comparisons of results from LC cases against control groups (HC for serum and urine or NCC for BALF) revealed the alteration of 26 annotated metabolites in serum and 32 in urine. Finally, BALF samples results from LC and NCC showed 16 metabolites altered by the disease which six metabolites were common in the three fluids: malonic acid, palmitic acid, phosphoric acid, inositol, isocitric acid and L-glycine being phosphoric acid the only metabolite with good sensitivity and specificity (AUC<0.75) for LC in the three fluids.

Moreover, the study allows establishing the following common altered pathways: glycine, serine and threonine metabolism, arginine and proline metabolism, inositol phosphate metabolism, alanine, aspartate and glutamate metabolism, pyruvate metabolism, galactose metabolism and cysteine and methionine metabolism, which are the most altered pathways in LC according to the corresponding p and Impact values.

Finally, it can be mentioned that several metabolite classes suffered alteration by the LC progression, such as amino acids, particularly L-glycine the only amino acid whose level decreased in the three fluids, which is in accordance with previous results in serum and plasma. Also, the levels of fatty acids such as palmitic, arachidonic, oleic and stearic acids were

altered in fluids from LC patients, as a consequence of proliferating tumor cells that use long chain fatty acids for membrane assembly and energy production. In addition, energy-related molecules were altered, such as sugars, and precursors as glycerol. At last, decreased levels of inositol, a component of cell membranes required by human cells for growth, were also observed in the three fluids from LC patients.

ACKNOWLEDGEMENTS

This work has been supported by the projects CTM2015-67902-C2-1-P from the Spanish Ministry of Economy and Competitiveness and P12-FQM-0442 from the Regional Ministry of Economy, Innovation, Science and Employment (Andalusian Government, Spain) and Neumosur scholarchip number 8/2012 and Separ scholarship number 124/2012. Belén Callejón Leblic thanks the Ministerio de Educación for a predoctoral scholarship FPU13/03615. Finally, authors are grateful to FEDER (European Community) for financial support, grants number UNHU13-1E-1611 and and UNHU15-CE-3140.

References

[1] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, et al., Cancer statistics, 2008., CA. Cancer J. Clin. 58 (2008) 71–96. doi:10.3322/CA.2007.0010.

[2] S. Hori, S. Nishiumi, K. Kobayashi, M. Shinohara, Y. Hatakeyama, Y. Kotani, et al., A metabolomic approach to lung cancer., Lung Cancer. 74 (2011) 284–92. doi:10.1016/j.lungcan.2011.02.008.

[3] R. Molina, J.M. Auge, J.M. Escudero, R. Marrades, N. Viñolas, E. Carcereny, et al., Mucins CA 125, CA 19.9, CA 15.3 and TAG-72.3 as tumor markers in patients with lung cancer: Comparison with CYFRA 21-1, CEA, SCC and NSE, Tumor Biol. 29 (2009). doi:10.1159/000181180.

[4] H.I. Pass, D.G. Beer, S. Joseph, P. Massion, Biomarkers and molecular testing for early detection, diagnosis, and therapeutic prediction of lung cancer, Thorac. Surg. Clin. 23 (2013). doi:10.1016/j.thorsurg.2013.01.002.

[5] L. Yu, K. Li, X. Zhang, Next-generation metabolomics in lung cancer diagnosis , treatment and precision medicine : mini review, Oncotarget. 8 (2017) 115774–115786.

[6] S. Bamji, S. Victor, D.M. Miller, H.B. Frieboes, A review of metabolism-associated biomarkers in lung cancer diagnosis and treatment, Metabolomics. 14 (2018) 1–16. doi:10.1007/s11306-018-1376-2.

[7] C.M. Rocha, A.S. Barros, B.J. Goodfellow, I.M. Carreira, A. Gomes, V. Sousa, et al., NMR metabolomics of human lung tumours reveals distinct metabolic signatures for adenocarcinoma and squamous cell carcinoma, Carcinogenesis. 36 (2015). doi:10.1093/carcin/bgu226.

[8] R. Ramautar, A. Demirci, G.J. de Jong, Capillary electrophoresis in metabolomics, TrAC Trends Anal. Chem. 25 (2006) 455–466. doi:10.1016/j.trac.2006.02.004.

[9] Y. Chen, Z. Ma, A. Li, H. Li, B. Wang, J. Zhong, et al., Metabolomic profiling of human serum in lung cancer patients using liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry and gas chromatography/mass spectrometry, J. Cancer Res. Clin. Oncol. 141 (2015). doi:10.1007/s00432-014-1846-5.

[10] M. Beckmann, D. Parker, D.P. Enot, E. Duval, J. Draper, High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry, Nat. Protoc. 3 (2008) 486–504. doi:10.1038/nprot.2007.500.

[11] Y. Chen, Z. Ma, A. Li, H. Li, B. Wang, J. Zhong, et al., Metabolomic profiling of human serum in lung cancer patients using liquid chromatography / hybrid quadrupole time-of-flight mass spectrometry and gas chromatography / mass spectrometry, (2015) 705–718. doi:10.1007/s00432-014-1846-5.

[12] L. Puchades-Carrasco, E. Jantus-Lewintre, C. Pérez-Rambla, F. García-García, R. Lucas, S. Calabuig, et al., Serum metabolomic profiling facilitates the non-invasive identification of metabolic biomarkers associated with the onset and progression of non-small cell lung cancer, Oncotarget. 7 (2016). doi:10.18632/oncotarget.7354.

[13] P.G. Lokhov, O.N. Kharybin, A.I. Archakov, Diagnosis of lung cancer based on direct-infusion electrospray mass spectrometry of blood plasma metabolites, Int. J. Mass Spectrom. 309 (2012) 200–205. doi:10.1016/j.ijms.2011.10.002.

[14] J. Maeda, M. Higashiyama, A. Imaizumi, T. Nakayama, H. Yamamoto, T. Daimon, et al., Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: a case control study., BMC Cancer. 10 (2010) 690. doi:10.1186/1471-2407-10-690.

[15] Y. Hanai, K. Shimono, K. Matsumura, A. Vachani, S. Albelda, K. Yamazaki, et al., Urinary volatile compounds as biomarkers for lung cancer., Biosci. Biotechnol. Biochem. 76 (2012) 679–84. doi:10.1271/bbb.110760.

[16] K. O'Shea, S.J.S. Cameron, K.E. Lewis, C. Lu, L.A.J. Mur, Metabolomic-based biomarker discovery for non-invasive lung cancer screening: A case study, Biochim. Biophys. Acta - Gen. Subj. 1860 (2016). doi:10.1016/j.bbagen.2016.07.007.

[17] C.M. Rocha, A.S. Barros, A.M. Gil, B.J. Goodfellow, E. Humpfer, M. Spraul, et al., Metabolic profiling of human lung cancer tissue by 1H high resolution magic angle

spinning (HRMAS) NMR spectroscopy., J. Proteome Res. 9 (2010) 319–32. doi:10.1021/pr9006574.

[18] A. Escribano Montaner, A. Moreno Galdó, Técnicas fibrobroncoscópicas especiales: lavado broncoalveolar, biopsia bronquial y biopsia transbronquial, An. Pediatría. 62 (2005) 352–366. doi:10.1157/13073249.

[19] W.E. Ho, Y.-J. Xu, C. Cheng, H.Y. Peh, S.R. Tannenbaum, W.S.F. Wong, et al., Metabolomics Reveals Inflammatory-Linked Pulmonary Metabolic Alterations in a Murine Model of House Dust Mite-Induced Allergic Asthma., J. Proteome Res. 13 (2014) 3771–3782. doi:10.1021/pr5003615.

[20] J.E. Wolak, C.R. Esther, T.M. O'Connell, Metabolomic analysis of bronchoalveolar lavage fluid from cystic fibrosis patients., Biomarkers. 14 (2009) 55–60. doi:10.1080/13547500802688194.

[21] B. Callejón-Leblic, T. García-Barrera, J. Grávalos-Guzmán, A. Pereira-Vega, J.L. Gómez-Ariza, Metabolic profiling of potential lung cancer biomarkers using bronchoalveolar lavage fluid and the integrated direct infusion/ gas chromatography mass spectrometry platform, J. Proteomics. 145 (2016) 197–206. doi:10.1016/j.jprot.2016.05.030.

[22] S.J. Bruce, P. Jonsson, H. Antti, O. Cloarec, J. Trygg, S.L. Marklund, et al., Evaluation of a protocol for metabolic profiling studies on human blood plasma by combined ultra-performance liquid chromatography/mass spectrometry: From extraction to data analysis., Anal. Biochem. 372 (2008) 237–49. doi:10.1016/j.ab.2007.09.037.

[23] M. Katajamaa, M. Oresic, Data processing for mass spectrometry-based metabolomics., J. Chromatogr. A. 1158 (2007) 318–28. doi:10.1016/j.chroma.2007.04.021.

[24] T. Wen, L. Gao, Z. Wen, C. Wu, C.S. Tan, W.Z. Toh, et al., Exploratory investigation of plasma metabolomics in human lung adenocarcinoma., Mol. Biosyst. 9 (2013) 2370–8. doi:10.1039/c3mb70138g.

[25] S.J.S. Cameron, K.E. Lewis, M. Beckmann, G.G. Allison, R. Ghosal, P.D. Lewis, et al., The metabolomic detection of lung cancer biomarkers in sputum, Lung Cancer. 94 (2016). doi:10.1016/j.lungcan.2016.02.006.

[26] W.C. Zhang, S.-C. Ng, H. Yang, A. Rai, S. Umashankar, S. Ma, et al., Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis, Cell. 148 (2012). doi:10.1016/j.cell.2011.11.050.

[27] A. Klupczynska, P. Dereziński, W. Dyszkiewicz, K. Pawlak, M. Kasprzyk, Z.J. Kokot, Evaluation of serum amino acid profiles' utility in non-small cell lung cancer detection in Polish population, Lung Cancer. 100 (2016). doi:10.1016/j.lungcan.2016.04.008.

[28] Y. Li, X. Song, X. Zhao, L. Zou, G. Xu, Serum metabolic profiling study of lung cancer using ultra high performance liquid chromatography/quadrupole time-of-flight mass spectrometry, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 966 (2014). doi:10.1016/j.jchromb.2014.04.047.

[29] D. Heber, R.T. Chlebowski, D.E. Ishibashi, J.N. Herrold, J.B. Block, Abnormalities in glucose and protein metabolism in noncachectic lung cancer patients, Cancer Res. 42 (1982) 4815–4819. http://www.scopus.com/inward/record.url?eid=2-s2.0-0020426260&partnerID=tZOtx3y1.

Figure Captions

Figure 1. PLS-DA score plots from a) serum b) urine and c) BALF from LC patients (red points) and controls, HC for serum and urine samples, or NCC for BALF samples (black points).

Figure 2. a) Chromatrogram of urine samples from LC patients without urease treatment, b) Chromatrogram of urine samples from LC patients with urease treatment.

Figure 3. Euler diagram of altered metabolites in LC from serum, urine and BALF samples.

Figure 4. Overview of the most important metabolomic changes observed in serum, urine and BALF from LC patients. a: Glycine, serine and threonine metabolism, b: Alanine, aspartate and glutamate, c: Inositol phosphate metabolism, d: Arginine and proline metabolism, e: Cysteine and methionine metabolism, f: TCA cycle, g: Galactose metabolism, h: Pyruvate metabolism. P value is the p calculated from the enrichment analysis and Impact is the pathway impact value calculated from pathway topology analysis.

Figure 5. Scheme of the most altered metabolic pathways in LC determined in the pathway analysis.