

Metabolic Signatures of Lung Cancer in Biofluids: NMR-Based Metabonomics of Blood Plasma

Cláudia M. Rocha,[†] Joana Carrola,[†] António S. Barros,[‡] Ana M. Gil,[†] Brian J. Goodfellow,^{†,⊥} Isabel M. Carreira,^{§,⊥} João Bernardo,^{#,⊥} Ana Gomes,^{#,⊥} Vitor Sousa,^{#,||,⊥} Lina Carvalho,^{#,||,⊥} and Iola F. Duarte^{*,†,⊥}

[†]CICECO and [‡]QOPNA, Department of Chemistry, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

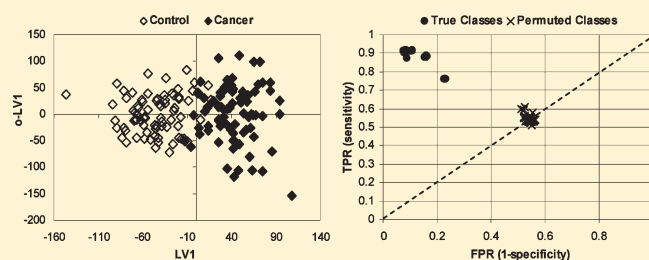
[§]Cytogenetics Laboratory and CNC, ^{||}Institute of Pathological Anatomy, and [⊥]CIMAGO, Faculty of Medicine, University of Coimbra, 3000 Coimbra, Portugal

[#]University Hospitals of Coimbra, 3000-075 Coimbra, Portugal

 Supporting Information

ABSTRACT: In this work, the variations in the metabolic profile of blood plasma from lung cancer patients and healthy controls were investigated through NMR-based metabonomics, to assess the potential of this approach for lung cancer screening and diagnosis. PLS-DA modeling of CPMG spectra from plasma, subjected to Monte Carlo Cross Validation, allowed cancer patients to be discriminated from controls with sensitivity and specificity levels of about 90%. Relatively lower HDL and higher VLDL + LDL in the patients' plasma, together with increased lactate and pyruvate and decreased levels of glucose, citrate, formate, acetate, several amino acids (alanine, glutamine, histidine, tyrosine, valine), and methanol, could be detected. These changes were found to be present at initial disease stages and could be related to known cancer biochemical hallmarks, such as enhanced glycolysis, glutaminolysis, and gluconeogenesis, together with suppressed Krebs cycle and reduced lipid catabolism, thus supporting the hypothesis of a systemic metabolic signature for lung cancer. Despite the possible confounding influence of age, smoking habits, and other uncontrolled factors, these results indicate that NMR-based metabonomics of blood plasma can be useful as a screening tool to identify suspicious cases for subsequent, more specific radiological tests, thus contributing to improved disease management.

KEYWORDS: lung cancer, NMR spectroscopy, metabonomics, blood plasma, metabolic profile



INTRODUCTION

Lung cancer is one of the most prevalent and fatal types of cancer, with average 5-year survival rates lower than 15%,^{1,2} which is mainly due to the advanced stage at which lung tumors are usually diagnosed. Indeed, when lung cancer is detected before metastasizing to lymph nodes or distant sites, the 5-year survival rates increase drastically to 60–80%, thus stressing the importance of early diagnosis. However, the majority of patients show no signs or symptoms during the initial phases of neoplastic growth, hindering early detection and the possibility of curative surgical treatment. Moreover, radiological tests, such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET), which would allow the detection of initial cancer lesions, are not suitable for general screening of the population, mainly due to their high costs. Therefore, new methods that can aid in the early detection of lung cancer and contribute to improved prognosis are greatly needed.

The search for metabolic markers of cancer in human tissues and biofluids has been the focus of a number of metabonomic studies in recent years.^{3,4} In particular, metabolic profiling of

blood plasma or serum has been increasingly used to unveil metabolic alterations associated with different cancer types, such as breast,^{5–7} kidney,^{8–10} liver,^{11–13} prostate,^{14–16} colorectal,^{17–20} oral,^{21,22} pancreatic,^{23,24} esophageal,²⁵ and bone²⁶ cancers. In the case of lung cancer, only a few studies focusing on plasma or serum metabolic profiling have been recently reported. Maeda and co-workers proposed that the differences in the plasma amino acid profiles between healthy controls and non-small-cell lung cancer (NSCLC) patients, as assessed by targeted liquid chromatography coupled to mass spectrometry (LC–MS), could potentially be useful for screening NSCLC.²⁷ In another MS study of specific compounds, namely, lysophosphatidylcholines (lysoPC), abnormal levels of lysoPC isomers with different fatty acyl positions were found in the plasma of lung cancer patients compared to controls.²⁸ Using a more global profiling approach, Jordan and colleagues reported the NMR analysis of paired tissue and serum samples from 14 subjects with two different lung cancer histological types (adenocarcinoma and squamous cell

Received: June 8, 2011

Table 1. Characteristics of Patients and Controls

	lung cancer patients	healthy controls
no. of subjects	85	78
gender	male 55 female 30	male 38 female 40
average age (range)	63 (30–85)	41 (22–60)
smoking habits	<i>a</i>	smokers 27 nonsmokers 38 ex-smokers 13
histopathology	adenocarcinoma 37 epidermoid carcinoma 18 carcinoid 11 sarcomatoid carcinoma 11 large cell carcinoma 5 small cell carcinoma 2 adenosquamous carcinoma 1	
TNM stage	I 46 II 23 III 7	

^aIn all cases, the surrounding parenchyma tissue presented histological characteristics of “smoking lung”.

carcinoma), as well as of serum from 7 healthy individuals.²⁹ This exploratory study, although involving a limited number of samples, showed the potential of serum NMR metabonomics to differentiate between the two lung cancer types, as well as between cancer patients and controls. In addition to cancer, plasma metabolic profiling has also been applied to investigate other lung disorders, namely, severe childhood pneumonia³⁰ and sepsis-induced acute lung injury (ALI),³¹ providing indication that distinct systemic metabolic signatures seem to be associated with different pulmonary pathologies.

Following our previous studies on tissues collected from surgical specimens^{32,33} and on urine from lung cancer patients and healthy subjects,³⁴ the present work aims to determine if the plasma metabolic composition reflects the presence of the disease and, thus, if nontargeted NMR-based profiling may be useful in lung cancer screening. Hence, preoperative plasma samples from patients with primary lung cancer ($n = 85$), not submitted to radiation or chemotherapy treatment, were analyzed by ¹H NMR, together with samples from healthy controls ($n = 78$), and their metabolic variations were assessed with the aid of multivariate statistical analysis. The discrimination ability between cancer and control groups was evaluated in terms of sensitivity and specificity levels through Monte Carlo Cross Validation (MCCV) applied to the PLS-DA models, and the differences in metabolite levels were further assessed through univariate analysis of signal integrals. Moreover, possible correlations to histological type and disease stage and the potentially confounding influence of gender, age, and smoking habits were investigated and discussed.

MATERIAL AND METHODS

Subjects

The characteristics of the subjects included in this study are summarized in Table 1. A total of 85 patients with primary lung cancer (30 female, 55 male, age range 30–85, average age 63) and 78 healthy control subjects (40 female, 38 male, age range 22–60, average age 41) were included in this study, following

informed consent and protocol approval by the University Hospitals of Coimbra. Final diagnosis and staging according to the TNM (tumor, nodes and metastases) system were established by the histopathological evaluation of surgical specimens. According to the 2004 WHO Classification of Tumours,³⁵ the malignant epithelial tumors characterized included 37 adenocarcinomas, 18 epidermoid carcinomas, 11 carcinoids, 11 sarcomatoid carcinomas, 5 large cell carcinomas, 2 small cell carcinomas, and 1 adenosquamous carcinoma. Regarding TNM staging, tumors were further classified as stage I ($n = 46$), stage II ($n = 23$), and stage III ($n = 7$). In all cases, the surrounding parenchyma tissue presented histological characteristics of “smoking lung”, indicating exposure to cigarette smoking (although actual smoking habits were not known). None of the patients had received radiation or chemotherapy treatment or shared any common medication regimen, although intake of other medicines (e.g., analgesics, antihypertensive drugs) could not be ruled out. Volunteers were included in the control group on the basis of a physicians’ assessment of their general health status (normal values in blood plasma and urine standard clinical tests, as well as absence of major illness or chronic medication). Among control subjects, 27 were smokers, 13 were ex-smokers (for more than 4 years), and 38 never smoked. Each individual (either patient or healthy volunteer) provided a blood sample after overnight fasting, to minimize dietary influence, although there was no control over previous food intake. Also, body mass index and nutritional status were not known. Blood was collected into a sodium-heparin tube and centrifuged ($1500 \times g$, 10 min) within a maximum of 30 min. Plasma aliquots of approximately 1 mL were then transferred into sterile cryovials, frozen, and stored at -80°C .

Sample Preparation

At the time of NMR analysis, plasma samples were thawed and homogenized using a vortex mixer. Then 400 μL of saline solution (NaCl 0.9% in 10% D₂O) was added to 200 μL of plasma. After centrifugation (8000 rpm, 5 min), 550 μL of each sample was transferred to 5 mm NMR tubes.

NMR Measurements

NMR spectra were acquired on a Bruker Avance DRX-500 spectrometer operating at 500.13 MHz for ¹H observation, at 300 K. For each sample, three 1D experiments were acquired: (a) standard ¹H spectrum with water suppression (“noesypr1d” in Bruker library, SW 10330.58 Hz, TD 32 K, relaxation delay 4 s, mixing time 100 ms, 128 scans); (b) T₂-edited (Carr-Purcell-Meiboom-Gill, CPMG) ¹H spectrum to allow better visualization of low molecular weight metabolites (“cpmgrp” in Bruker library, SW 10330.58 Hz, TD 32 K, relaxation delay 4 s, total spin-echo time 64 ms, 256 scans); and (c) diffusion-edited ¹H spectrum to select signals of bound or large molecules (“ledbpgp2s1dpr” in Bruker library, SW 10330.58 Hz, TD 32 K, relaxation delay 4 s, diffusion delay 100 ms, pulsed-field gradient 1 ms, 128 scans). All 1D spectra were processed with a line broadening of 0.3 Hz (standard and T₂-edited) or 0.5 Hz (diffusion-edited) and a zero filling factor of 2, manually phased and baseline corrected. The chemical shifts were referenced internally to the α -glucose signal at δ 5.23. 2D homonuclear (¹H–¹H TOCSY and *J*-resolved) and heteronuclear (¹H–¹³C HSQC) spectra were also registered for selected samples to aid spectral assignment, which was based on matching 1D and 2D data to reference spectra in the BBIREFCODE-2-0-0 database (Bruker Biospin, Rheinstetten, Germany), as well as other existing databases³⁶ and literature reports.

Multivariate Statistical Analysis

Multivariate analysis was applied to a total of 163 spectra from 85 cancer samples and 78 control plasmas. For each experiment type (standard 1D, CPMG, and diffusion-edited), data matrices were built in Amix version 3.9 (Bruker BioSpin, Rheinstetten, Germany) using all intensity values in the region δ 11 to -3 of the spectra (full resolution data, 44419 variables), excluding the subregion δ 5.15–4.5 to remove the variability arising from water suppression. After normalization (by adjusting the total area to unity), in order to minimize possible differences in concentration between samples, and scaling to unit variance, so that all variables were given equal weight regardless of their absolute value, Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), and Orthogonal Projections to Latent Structures (OPLS-DA) were performed in SIMCA-P11.5 (Umetrics, Umeå, Sweden). A default 7-fold internal cross validation was employed, from which Q^2 and R^2 values (representing, respectively, the predictive capability and the explained variance) were extracted. PCA, used to detect intrinsic clusters and outliers within the data set, was followed by (O)PLS-DA to maximize class discrimination. OPLS-DA is an extension of PLS-DA featuring an integrated Orthogonal Signal Correction (OSC) filter to remove variability not relevant to class separation.³⁷ Model robustness was assessed by Monte Carlo Cross Validation (MCCV)³⁸ with 500 iterations. In this procedure, the data were divided into a calibration set comprising 60% of samples and a validation set comprising the remaining 40% of samples. Then, each iteration, with randomly changed composition of the calibration and validation sets, performed internal cross validation of the calibration set using seven blocks and prediction of class membership for samples in the validation set. For each of the 500 randomly generated classification models, the number of Latent Variables (LV), the Q^2 (expressing the cross-validated explained variability), and the classification rate (percentage of samples correctly classified) were recovered. The selection of model complexity was based on the most frequent list of model properties that maximizes the classification rate (i.e., lower LV and higher Q^2). The sensitivity and the specificity of the model were then depicted from the confusion matrix resulting from the 500 MCCV iterations. Further validation was performed by permutation tests, consisting of randomly permuting class membership and running 500 MCCV iterations. The same multivariate statistical approach was applied to detect eventual correlations to histological types and disease stages and to assess the influence of other factors, such as gender, age, and smoking habits, by building new models according to these classifiers and verifying their discriminating power.

Quantitative Analysis of Variations in Selected Metabolites

Due to improved baseline and metabolite detection in the CPMG spectra, these were chosen for quantitative measurements by spectral integration, assuming that any reduction in signal intensity due to relaxation effects would be consistent across samples, thus not affecting the evaluation of relative changes in metabolite levels. Integrals of selected signals, normalized to total spectral area (δ 11 to -3 excluding δ 5.15–4.5), were compared between groups (e.g., control vs cancer), and the statistical significance between average values was determined using the two-sample *t* test or the nonparametric analogue Wilcoxon rank sum test with continuity correction. The boxplot representation was used to visualize the variation in the levels of integrated compounds in control and cancer samples.

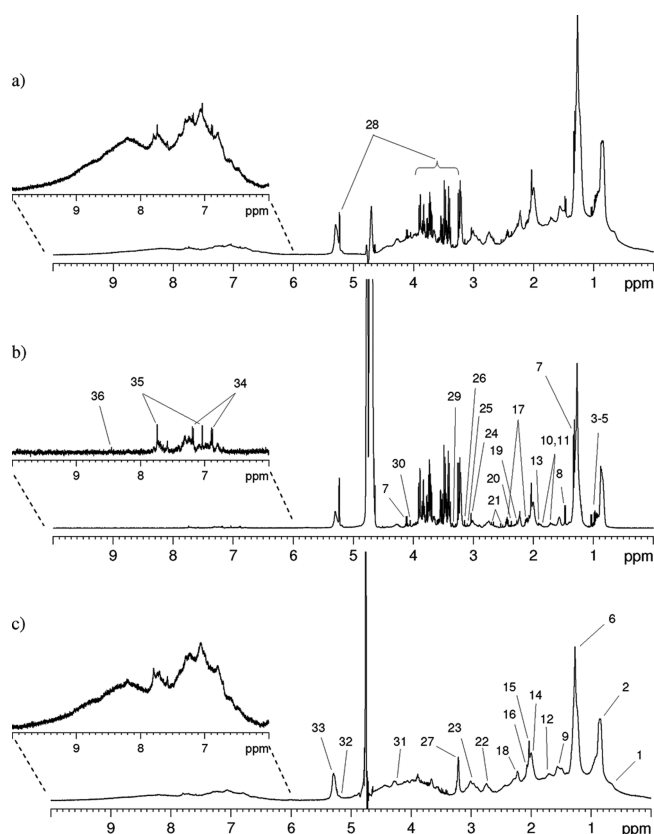


Figure 1. 500 MHz ^1H NMR spectra of blood plasma from a lung cancer patient: (a) standard 1D, (b) T_2 -edited (CPMG), (c) diffusion-edited. The low-field region (δ 6–10) is vertically scaled 8 times relative to the rest of the spectrum. Signal assignment: 1: cholesterol C18, 2: lipid CH_3 , 3: isoleucine, 4: leucine, 5: valine, 6: lipid $(\text{CH}_2)_m$, 7: lactate, 8: alanine, 9: lipid $\text{CH}_2\text{CH}_2\text{CO}$, 10: arginine, 11: lysine, 12: lipid $\text{CH}_2\text{CH}_2\text{C}=\text{C}$, 13: acetate, 14: lipid $\text{CH}_2\text{CH}=\text{CH}$, 15: NAC1, glycoprotein N-acetyl group NHCOCH_3 , 16: NAC2, glycoprotein N-acetyl group NHCOCCH_3 , 17: glutamine, 18: lipid CH_2CO , 19: unknown 2.27 ppm, 20: pyruvate, 21: citrate, 22: lipid $\text{CH}=\text{CHCH}_2\text{CH}=\text{CH}$, 23: albumin lysyl $\epsilon\text{-CH}_2$, 24: creatine, 25: unknown 3.10 ppm, 26: unknown 3.14 ppm, 27: phospholipids choline headgroup $-\text{N}^+(\text{CH}_3)_3$, 28: glucose, 29: methanol, 30: creatinine, 31: glyceryl of lipids CH_2OCOR , 32: glyceryl of lipids CHOCOR , 33: unsaturated lipid $-\text{CH}=\text{CH}-$, 34: tyrosine, 35: histidine, 36: formate.

RESULTS

Discrimination between Control and Lung Cancer Patients

The standard ^1H 1D spectrum of plasma (Figure 1a) showed broad resonances from proteins (e.g., albumin) and lipoproteins, superimposed with narrow signals from low molecular weight metabolites. By using a CPMG sequence, the broad protein envelope was clearly attenuated, significantly improving the visibility of metabolite signals, particularly in the low and high frequency regions (Figure 1b). Assigned metabolites included several amino acids (e.g., valine, alanine, lysine, glutamine, tyrosine, histidine), organic acids (e.g., lactate, acetate, acetoacetate, pyruvate, citrate, formate), creatine, creatinine, methanol (confirmed by spiking experiments), and glucose. The resonances from fatty acyl chain protons in lipoproteins were also clearly visible in the CPMG spectrum, although with attenuated intensities relative to the standard 1D spectrum. On the other hand, the diffusion-edited

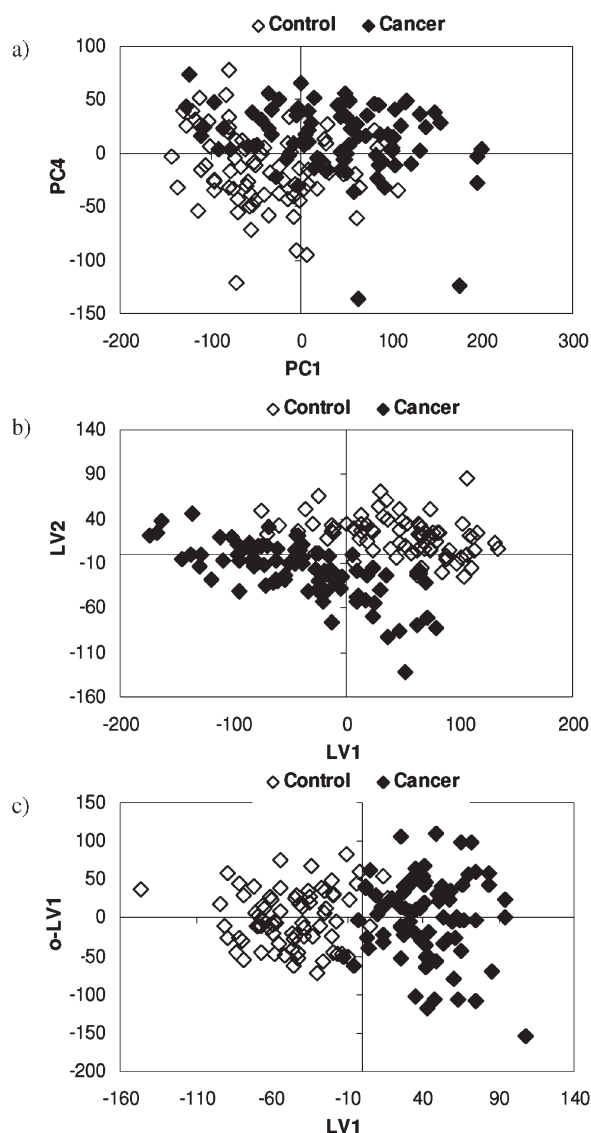


Figure 2. Scores scatter plots (a) PC1 vs PC4, (b) LV1 vs LV2, (c) LV1 vs o-LV1 resulting from applying, respectively, PCA, PLS-DA (LV 2, R^2X 0.17, R^2Y 0.73, Q^2 0.57), and OPLS-DA (LV 1 + 1, R^2X 0.17, R^2Y 0.73, Q^2 0.54) to the 1H CPMG NMR spectra of plasma from healthy controls (\diamond) and lung cancer patients (\blacklozenge).

spectrum (Figure 1c) showed the macromolecular profile comprising plasma proteins and lipids.

Visual inspection of average plasma spectra from control and cancer subjects (Supplementary Figure S1) suggested differences in their lipoprotein profiles and in the relative amount of some metabolites (e.g., lactate, histidine). However, confirmation of these and other subtle differences requires application of multivariate analysis. Figure 2 shows the scores scatter plots obtained by PCA, PLS-DA, and OPLS-DA applied to the CPMG plasma spectra. A trend for unsupervised separation between control and cancer samples was found in the PC1 vs PC4 scores plot (Figure 2a). PLS-DA maximized group separation (Figure 2b), which was further improved along LV1 when OPLS-DA was employed (Figure 2c).

To better assess the predictive ability of the classification models built for the three experiment types, Monte Carlo Cross

Table 2. Average Prediction Results Obtained by MCCV (500 Iterations) of PLS-DA Models Based on Standard 1D, CPMG, and Diffusion-Edited Spectra of Plasma from Healthy Controls ($n = 78$) and Lung Cancer Patients ($n = 85$)^a

	median Q^2	sensitivity (%)	specificity (%)	classification rate (%)
standard 1D	0.58	78.2	89.6	83.6
CPMG	0.64	91.5	89.2	90.4
diffusion-edited	0.62	83.9	89.4	86.5

^aSensitivity was calculated from the ratio of true positives (cancer samples correctly predicted) to total number of modeled cancer spectra, whereas specificity was determined from the ratio of true negatives (control samples correctly predicted) to total number of modeled control spectra.

Validation (MCCV) was applied, as described in Material and Methods. The results are summarized in Table 2 in terms of Q^2 values, expressing the models' predictive power, sensitivity (% of cancer samples correctly classified – true positives), specificity (% of control samples correctly classified – true negatives), and the classification rate (total number of samples correctly classified). Overall, the best results were obtained for the model based on the CPMG spectra, with a median Q^2 of 0.64, sensitivity and specificity levels of 91.5% and 89.2%, respectively, and a 90.4% classification rate. Further model validation relied on permutation analysis, also employing 500 MCCV iterations, with randomly permuted class membership. The prediction results obtained for the real (with true classes assigned) and the permuted models were plotted in the Receiver Operating Characteristic (ROC) space (Figure 3a). While the real model showed high sensitivity and specificity for most iterations, the permuted model results fell along the diagonal of the ROC space, i.e., the line of no discrimination. Also, the Q^2 values were negative for most permuted models (Figure 3b), and the classification rate decreased to 49.4%, thus validating the predictive ability of the real model.

To identify the variations responsible for class discrimination and facilitate their interpretation, the OPLS-DA LV1 loadings colored according to Variable Importance in Projection (VIP) were carefully inspected, and when possible, the signals showing high VIP values (>1.0) were assigned to specific compounds (Figure 4). A major feature highlighted by these loadings was the different proportion of the main lipoprotein subclasses in the plasma samples of control and cancer subjects, as shown by the splitting of the main fatty acyl chain resonances into positive and negative loading signals. In particular, negative loadings were found for broad resonances at δ 0.83, 1.23, and 5.26, whereas the signals at δ 0.87, 1.27, and 5.30 showed positive loadings. On the basis of the literature^{39–41} and the LC-NMR data collected in our laboratory for other human plasma samples, these groups of resonances correspond mainly to HDL and LDL + VLDL lipoproteins, respectively, thus suggesting that the former were decreased in cancer, whereas the latter were increased relative to controls. Moreover, the negative loading signal for the δ 3.21 resonance, arising from methyl head groups of choline-containing phospholipids, further supports the assignment to HDL, since phosphatidylcholine is known to be the most predominant lipid in the HDL fraction.⁴¹ Also of note is the positive intensity of the signals arising from the glyceryl backbone of lipids, with clearly high VIP values in the loadings calculated from standard

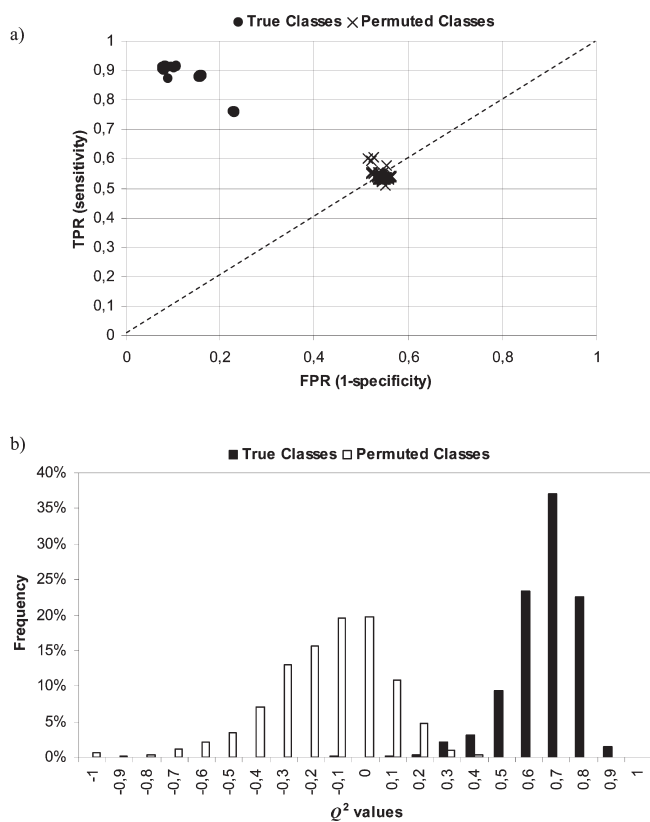


Figure 3. (a) Receiver Operating Characteristic (ROC) space, where each point represents a prediction result (sensitivity and 1-specificity) of the confusion matrices obtained from MCCV (500 iterations) of the PLS-DA model based on ^1H CPMG plasma spectra; (●) prediction results in the original model (class membership correctly assigned); (×) prediction results in the permuted model (class membership randomly permuted). (b) Distribution of Q^2 values obtained by MCCV (500 iterations) of the same PLS-DA models (original and permuted).

1D and diffusion-edited spectra (not shown), suggesting the total amount of triglycerides to be higher in the plasma of cancer patients.

In regard to small metabolites, the OPLS-DA loadings showed clearly positive signals with high VIP values (>1.0) for lactate and pyruvate, suggesting that these acids were increased in cancer samples, and signals with negative intensity and high VIPs for some amino acids (valine, alanine, lysine/arginine, glutamine, tyrosine, and histidine), acetate, citrate, formate, glucose, and methanol, suggesting their levels to be lower in the plasma of cancer patients compared to controls. Moreover, a number of other signals in the low and high frequency regions (indicated with asterisks in Figure 4) seemed to be important for sample discrimination, but in part due to their low intensities, their identity remains unknown at this time, thus precluding further interpretation of their variations.

To verify the variations highlighted by loadings inspection, specific signals in the CPMG spectra have been integrated, and the mean areas between control and cancer groups have been compared. With the exception of the signals from lipoprotein subclasses and lysine/arginine, which could not be integrated because of severe spectral overlap, all other assigned signals showing high VIP values were found to present statistically significant differences (p -value <0.001) between control and cancer

plasma samples. Table 3 summarizes the discriminating metabolites and their percentage variations in cancer samples relative to control plasma. Supplementary Figures S2 and S3 show the boxplot representations for the relative levels of these compounds.

Discrimination According to Histological Type and Stage

The possible correlation of plasma profiles to tumor histological type was investigated. In a first approach, the four main types, adenocarcinomas ($n = 32$), epidermoid carcinomas ($n = 18$), carcinoids ($n = 11$), and sarcomatoid carcinomas ($n = 11$), were considered together. However, no type-related clustering could be observed in PCA, and PLS-DA produced no valid discrimination between the four classes considered. Then, pairwise modeling was performed, including two histological types at a time. Interestingly, plasma samples from patients with carcinoid and sarcomatoid tumors could be reasonably discriminated by PLS-DA, especially when considering the diffusion-edited spectra. Figure 5a and b show, respectively, the scores scatter plots resulting from PLS-DA and OPLS-DA modeling of these spectra, where sample separation along LV1 was clear. The OPLS-DA LV1 loadings (Figure 5c) suggested that the plasma from carcinoid patients was characterized by relatively higher amounts of HDL lipoproteins (as indicated by the chemical shifts of negative red-colored loading signals), whereas the plasma from patients with sarcomatoid carcinomas was richer in glycoproteins, indicated by signals at δ 2.03 and 2.07 (arising from *N*-acetyl groups of *N*-acetylated carbohydrate side chains) and at δ 3.3–4.1 (arising from carbohydrate ring protons).⁴² The same metabolic differences were found when considering the standard 1D spectra, whereas modeling of CPMG spectra produced no valid discrimination between the two histological types. By MCCV with 500 iterations, PLS-DA of the diffusion-edited spectra produced a moderate classification rate of 77.6%, with a median Q^2 value of 0.62. Clearly, more samples of each type are needed to increase the robustness of this classification and confirm the metabolic differences highlighted in this preliminary analysis.

Regarding stage, when considering only the plasma samples from patients with cancer at stages I ($n = 46$), II ($n = 23$), and III ($n = 7$), excluding the control group, multivariate modeling produced no valid discrimination between those classes, thus suggesting a weak ability of plasma profile to predict disease stage. However, each stage could be discriminated from the control group, confirming the results obtained for the whole data set. Most importantly, plasma from patients with stage I cancer could still be differentiated from control plasma, with a classification rate of 89.0% and a reasonable predictive ability, expressed by a median Q^2 value of 0.56. The PLS-DA model specificity, as determined by MCCV, was 92.1%, whereas the sensitivity decreased to 83.1%, expressing a higher false negative rate (higher number of cancer samples misclassified as healthy) compared to the model comprising all samples. If, on one hand, this decrease in sensitivity may reflect the lower impact of stage I tumors on plasma composition, the lower number of samples included in the cancer class ($n = 46$ vs total $n = 85$) may also account for this result. In fact, the OPLS-DA loadings explaining the separation between control and stage I cancer samples (Supplementary Figure S4) highlighted the same metabolic differences as when including all the plasma samples in the cancer class (regardless of stage), thus showing that these differences were present in the plasma, right from the initial phases of tumor development.

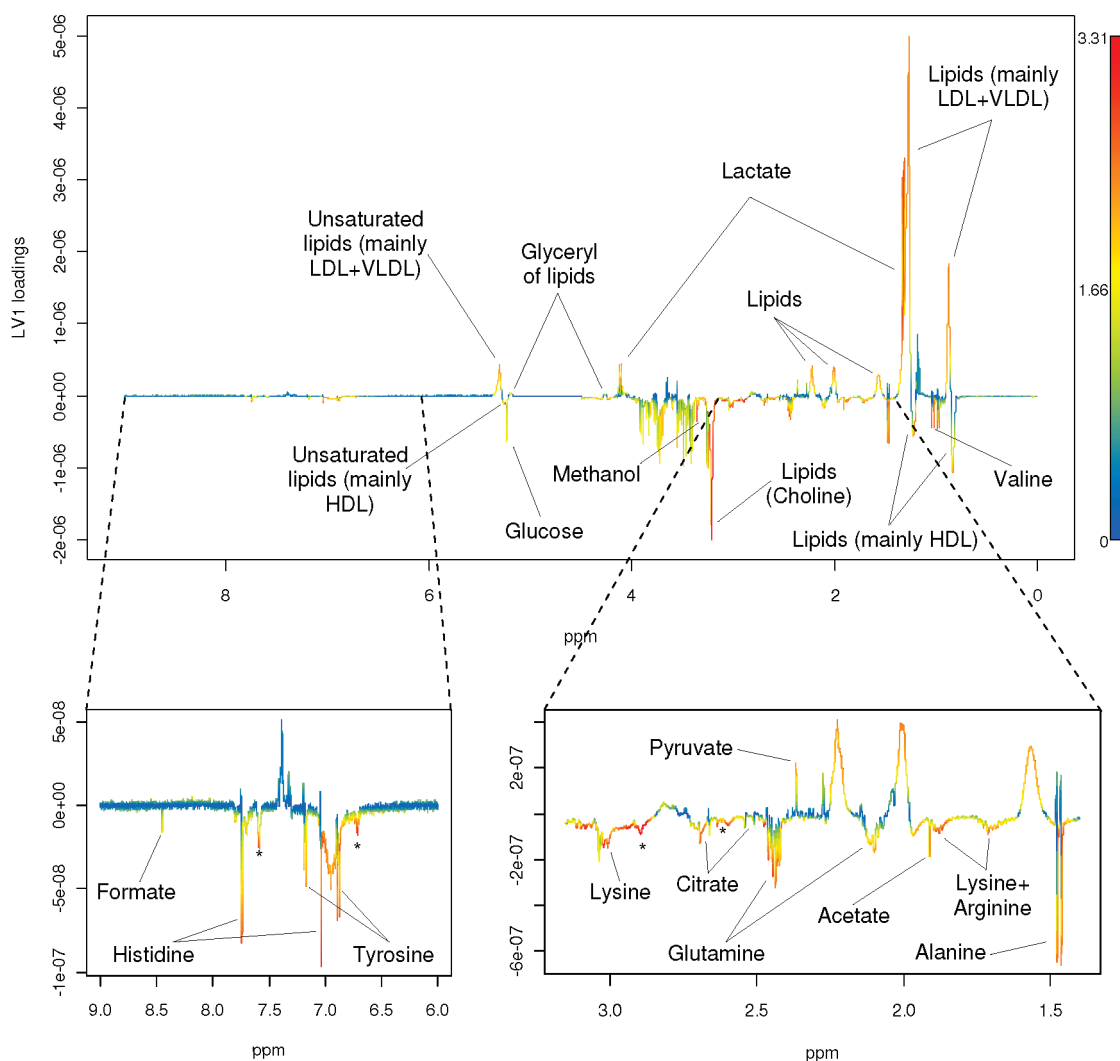


Figure 4. OPLS-DA LV1 loadings plot colored as a function of VIP. Assignment of main signals is indicated (unassigned signals with high VIP are marked with an asterisk).

Modeling of Other Possible Influencing Factors (Gender, Age, Smoking Habits)

Although the metabolic composition of plasma is relatively stable, a number of extrinsic and intrinsic factors may account for some interindividual variability and influence the interpretation of disease-related effects. Therefore, the contribution of some of these factors, namely, gender, age, and smoking habits, to the plasma profiles measured in this study were assessed by multivariate analysis of CPMG spectra (found to provide the best discrimination between control and cancer samples).

Both control and cancer groups comprised a balanced number of males and females, thus decreasing the probability of gender-related bias. Indeed, when the PCA scores scatter plots corresponding to all samples were colored according to gender, no clear gender-related clustering was observed. Still, in order to compare the discriminatory power of gender with that of the disease itself, the data set has been modeled by PLS-DA (with 500 MCCV iterations) taking the gender as classifier. Much lower Q^2 value and classification rate were obtained for this model (Table 4). These results show that the metabolic differences arising from gender were not as important for sample discrimination as those related to the presence of the disease. Also, the

loadings explaining the separation between genders (Supplementary Figure S5) revealed that the main differences between males and females were in the lipoprotein profile (HDL increased in females) and in the levels of creatine (increased in females, p -value 9.48×10^{-6}) and creatinine (increased in males, p -value 6.15×10^{-12}). Lactate and histidine signals also presented relatively high VIP values, but the difference in their integrals between the two genders did not reach statistical significance. Therefore, with the exception of lipoproteins, which also varied between control and cancer samples, distinct metabolites accounted for gender- and disease-related effects.

In regard to the variability relating to age, this study had an important limitation, which was the large difference in the age distributions of control and cancer groups. This age difference was due, on the one hand, to the greater prevalence of lung cancer on older subjects and, on the other hand, to the difficulty in finding healthy volunteers of more advanced ages. To address this problem and assess the potentially confounding influence of age on the discrimination between controls and patients, the following strategy has been followed: First, the PCA scores scatter plots corresponding to the whole data set ($n = 163$) were colored according to five age ranges (≤ 40 , 41–50, 51–60, 61–70, ≥ 71).

No age-dependent clustering or trend was observed in any of the PC scores plots (not shown). Then, a subgroup of subjects with

Table 3. Main Metabolites Contributing to the Discrimination between Plasma of Lung Cancer Patients and of Healthy Subjects

metabolite	$\delta^1\text{H}$ (ppm)	% average change (relative to controls) ^a	<i>p</i> -value
acetate	1.91	-16.7 ± 3.1	2.81×10^{-8}
alanine	1.47	-12.81 ± 2.4	3.37×10^{-8}
citrate	2.52	-11.0 ± 4.3	1.36×10^{-4}
formate	8.45	-30.5 ± 9.2	3.66×10^{-5}
glucose	3.52	-11.2 ± 2.6	8.46×10^{-6}
glutamine	2.44	-20.2 ± 2.9	1.04×10^{-12}
histidine	7.04	-26.6 ± 2.8	2.20×10^{-16}
lactate	4.11	18.8 ± 4.0	3.51×10^{-5}
methanol	3.35	-33.8 ± 4.2	6.66×10^{-16}
pyruvate	2.36	16.7 ± 3.6	1.84×10^{-5}
tyrosine	7.18	-10.8 ± 3.0	2.14×10^{-4}
valine	1.03	-14.4 ± 2.7	5.04×10^{-8}

^a Average changes determined by spectral integration of the signal specified in the second column. *P*-values computed by comparison of mean integrals between the two groups, as described in Material and Methods.

ages between 41 and 60 ($n = 72$) was considered, to include a balanced number of controls ($n = 41$, average age 49 ± 5) and patients ($n = 31$, average age 54 ± 5). Compared to the whole data set ($n = 163$), the discrimination between control and cancer samples degraded when this age subgroup was considered, as evaluated by the lower predictive ability (median Q^2 0.45) and classification rate (81.6%) of the resulting model (Table 4). On the other hand, when the classifier was age, attributing different classes to subjects in the age ranges 41–50 and 51–60, the PLS-DA model showed no predictive power at all, with negative Q^2 and low classification rate (Table 4). Therefore, in the subgroup considered, the presence of the disease was confirmed to have a more important influence in the plasma profiles than the subject's age.

In respect to the influence of smoking habits, the heterogeneous nature of the control group, comprising smokers ($n = 27$), nonsmokers ($n = 38$), and former smokers ($n = 13$), is expected to reduce possible bias in the discrimination between controls and cancer patients (the pulmonary parenchyma tissue of which presented histological patterns of “smoking lung”). Still, in order to better assess the potentially confounding influence of cigarette smoke, subgroups of control subjects, defined according to their smoking habits, have been modeled by PLS-DA together with the cancer class. The statistical results obtained are presented in Table 4. When only smokers were included in the control group

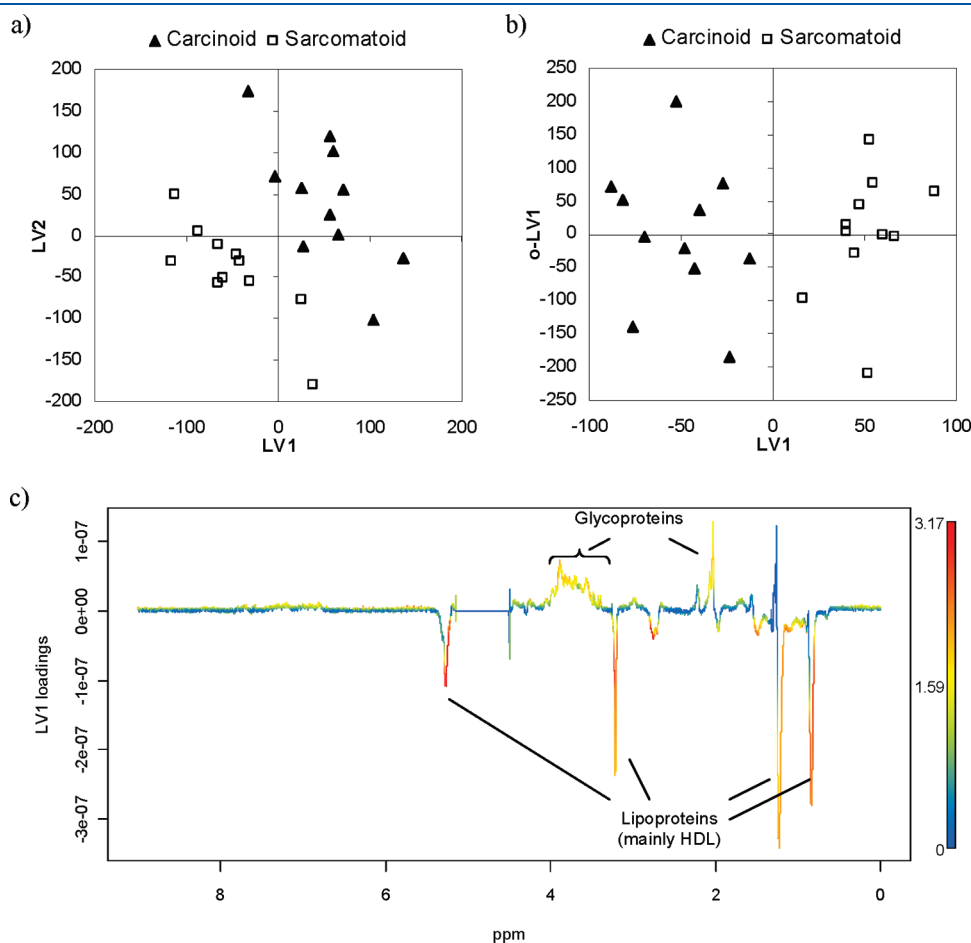


Figure 5. (a) LV1 vs LV2 scores scatter plot resulting from applying PLS-DA to the diffusion-edited spectra of plasma from patients with carcinoid ($n = 11$) and sarcomatoid tumors ($n = 11$) (LV 2, R^2X 0.35, R^2Y 0.86, Q^2 0.54). (b) LV1 vs o-LV1 scores scatter plot resulting from applying OPLS-DA to the same data (LV 1 + 1, R^2X 0.35, R^2Y 0.86, Q^2 0.37). (c) OPLS-DA LV1 loadings plot colored as a function of VIP.

Table 4. Prediction Results Obtained by MCCV (500 Iterations) of PLS-DA Models Based on CPMG Plasma Spectra^a

	median Q^2	sensitivity (%)	specificity (%)	classification rate (%)
control (78) vs cancer (85)	0.64	91.5	89.2	90.4
male (93) vs female (70)	0.36	63.2	76.6	70.9
control (41) vs cancer (31), age 41–60	0.45	70.4	90.2	81.6
age 41–50 (33) vs 51–60 (39)	−0.095	60.3	52.7	56.8
control smokers (27) vs cancer (85)	0.47	96.7	51.8	85.9
control nonsmokers (38) vs cancer (85)	0.47	94.7	80.0	90.2
control nonsmokers (38) vs control smokers (27)	−0.072	36.5	69.6	55.8

^aData for the following classes: control vs cancer; male vs female; control vs cancer (sub-group of subjects in the 41–60 age range); age 41–50 vs 51–60; control vs cancer varying the composition of the control group according to smoking habits; control nonsmokers vs control smokers. Sample numbers for each class are presented within brackets in the first column.

($n = 27$), a lower predictive power resulted (Q^2 0.47) and the specificity decreased drastically to 51.8%, meaning that more false positives (controls misclassified as cancer patients) arose from the model. Still, the overall classification rate remained relatively high (85.9%). On the other hand, when only nonsmokers were included in the control group ($n = 38$), class discrimination was not improved relative to taking all controls regardless of their smoking habits. Q^2 was close to 0.5 and the classification rate remained at about 90%. Moreover, the loadings profiles resulting from these models (not shown) were identical to those explaining the discrimination between control and cancer samples in the total model. To further assess the specific impact of smoke on the plasma profiles, control samples, divided into smokers and nonsmokers, have also been taken for PLS-DA modeling. As shown in Table 4, a negative median Q^2 and low classification rate (55.8%) were obtained, clearly showing that the plasma profiles from healthy nonsmokers could not be discriminated from those of control smokers, thus hindering any further speculation about putative plasma metabolites related to smoking habits.

DISCUSSION

An important feature accounting for the discrimination between control and cancer samples was the difference in the relative amounts of the major lipoprotein subclasses: HDL was lower in the plasma of patients, whereas LDL + VLDL and total triglycerides were increased relative to healthy subjects. These results agree with previous studies reporting significant alterations in the serum lipoprotein profile in a variety of hematological and solid tumors, including lung cancer,^{43,44} and thus are probably not specific to this type of cancer. In particular, low HDL was found to be the most discriminating factor in a study comparing over 500 patients with various types of solid tumors and over 900 age- and gender-matched healthy controls.⁴⁴ Other commonly observed alterations comprise the increase in triglycerides, also apparent in our study, and the reduction in LDL. This latter alteration, however, could not be observed in the present work as the contribution of LDL could not be evaluated separately from that of VLDL due to spectral overlap.

Although the biochemical origin of these changes is not well understood, it is possible that they relate to increased uptake of circulating lipoproteins to supply the necessary cholesterol for membrane build-up by fast-growing tumor cells. In fact, we have observed a trend for lipids to be increased in lung tumoral tissue (relative to noninvolved pulmonary parenchyma),^{32,33} and although this increase could have multiple sources, including

tissue necrosis, its possible relation with the observed changes in the plasma lipoprotein profile should not be ruled out. The poor nutritional status generally associated with cancer has also been reported to account for the reduction in HDL and LDL levels.⁴⁵ However, although the nutritional status has not been assessed in the present study, the increased levels of plasma triglycerides in cancer patients suggests that this was not a dominant factor. In fact, previous studies have made a similar observation and concluded that the nutritional status did not entirely explain the disturbances in lipoprotein composition.⁴³ Moreover, cardiovascular and inflammatory diseases have been ruled out as main causes for these disturbances by Muntoni et al.,⁴⁴ who suggested that the lipoprotein profile may have significance in the diagnosis and prognosis of solid tumors.

The plasma NMR profile of healthy subjects and cancer patients also differed in the relative amounts of a number of small molecules. Lactate was found to be increased and glucose was found to be decreased in the plasma of cancer patients, most likely reflecting the well-known increased glycolytic activity of tumor cells (the Warburg effect), as confirmed by metabolic profiling of lung tissues.^{32,33,46} Indeed, the standard measurement of glucose in serum samples from 300 fasted patients with a variety of cancer types also showed a significant reduction of glucose levels relative to the same number of gender- and age-matched controls.⁴⁴ Consistently with this enhanced glycolysis and as previously reported for leukemia,⁴⁷ kidney,⁸ liver,¹² and colorectal¹⁸ cancers, the glycolytic product pyruvate was found to be increased. Together with the observed reduced level of citrate, these results suggest a decreased Krebs cycle, in agreement with the commonly observed mitochondrial dysfunction in cancer.^{48,49}

The levels of a number of amino acids (alanine, glutamine, valine, tyrosine, and histidine) were found to be significantly decreased in the plasma of cancer patients relative to healthy controls. In contrast with these results, Maeda and coauthors²⁷ have recently reported significantly higher concentrations (p -value ≤ 0.001) of alanine, tyrosine, and several other amino acids (proline, glycine, isoleucine, phenylalanine, ornithine, and lysine) in the plasma of lung cancer patients and no significant difference for glutamine or valine, whereas histidine was the only amino acid found at lower concentrations than in controls. Although this discrepancy is intriguing, several aspects may account for it, such as the different genetic background and dietary/lifestyle habits of European subjects (enrolled in our study) and Japanese subjects (enrolled in the other work). Moreover, disease stage may also have an influence, since the plasma free amino acid profile has been shown to vary with cancer

progression.⁵⁰ Whereas in Maeda's study nearly half of the patients had cancer at advanced stage (III or IV), the large majority of our patients were at more initial stages (I or II). Considering other literature reports, plasma levels of alanine (and, in fact, of several gluconeogenic amino acids) have been found to be decreased in various cancer types,⁵⁰ including lung cancer,^{51,52} most probably due to its increased utilization as a major gluconeogenic precursor, to meet the high glucose uptake and demand by tumor cells. Additionally, the decrease in glutamine found, for instance, in the plasma of renal,^{8,9} pancreatic,²³ and other cancer patients⁵⁰ has been associated with increased glutaminolysis to help sustain the Krebs cycle and energy regeneration and to provide precursors for nucleic acid synthesis. Overall, as reviewed by Lai et al.,⁵⁰ despite the different specific amino acids varying for different cancer types and some inconsistencies between reports on the same cancer type, the large majority of studies reported decreased levels of most plasma free amino acids (PFAA) in cancer compared to healthy controls. Moreover, it has been shown that, whereas at advanced stages the altered PFAA is strongly correlated to malnutrition and cancer-associated weight loss,⁵³ the amino acid imbalance also occurs at early stages in nonmalnourished and noncachectic patients.⁵⁴

In addition to the above-mentioned metabolite changes, we found reduced levels of acetate, formate, and methanol in the plasma of lung cancer patients compared to controls. Acetate is an end product of lipid metabolism, and its decrease may reflect a reduced lipid catabolism in order to sustain accelerated cell proliferation. This has also been observed in the plasma of renal cell carcinoma patients⁹ and in tumor lung tissues compared to normal pulmonary parenchyma.^{32,33} Endogenous methanol has been previously reported in the blood plasma of healthy individuals,⁵⁵ but we found no previous record of its variation (or that of formate) in cancer patients. Actually, no explanation for the decreases in the levels of these two metabolites is proposed at this stage.

The influence of possible confounders, namely, gender, age, and smoking habits, on the discrimination between control and cancer samples has also been assessed. Gender was found to have a much lower discriminatory power than the disease, and the few gender-related metabolic differences encountered were in accordance with previously reported differences between males and females, namely, the relative abundance of lipoproteins subclasses^{56,57} and the levels of creatine and creatinine.³⁶ In regard to age, it was found to have a much smaller influence on the plasma composition than the disease, in line with the results reported by others for different cancer types, such as leukemia⁴⁷ or esophageal cancer.²⁵ However, possible bias arising from the disparate age distributions in controls (ages 22–60) and cancer subjects (ages 30–85) could not be ruled out, since the discrimination between these classes was found to be worse for a subgroup of subjects with more closely matched ages. Therefore, in the continuity of this work, a larger number of samples from older healthy individuals must be collected in order to improve age-matching between controls and patients and further assess the influence of age on sample discrimination.

Regarding the potentially confounding influence of smoking habits, a drastic decrease in model specificity was observed when only healthy smokers were included in the control group. However, the much smaller number of samples ($n = 27$) in this new control group could also account for this poorer predictive

ability. Indeed, if smoking was a major factor in control vs cancer discrimination, one would expect this discrimination to be enhanced when only nonsmokers were included in the control group and this was clearly not the case. Moreover, the metabolic features discriminating cancer from control subjects were the same irrespectively of the latter being smokers or not.

In regard to other possible confounders, although we tried to minimize the dietary influence by collecting all samples after overnight fasting, it cannot be completely excluded, as the metabolism of some food constituents can affect blood composition many hours after intake. However, as neither patients nor healthy subjects were under any controlled or standardized diet, it is unlikely that diet would account for specific differentiating effects between the two groups. Future studies should however entail the assessment of body mass index or the occurrence of weight loss or malnutrition in the patients group, as these might influence the metabolic variations, as discussed above for lipoproteins and amino acids.

CONCLUSIONS

This work has shown that NMR-based metabonomics of blood plasma is able to differentiate patients with primary lung cancer from healthy subjects with high sensitivity and specificity. Despite the expected interindividual metabolic variability and possible confounders (age, smoking habits, and other uncontrolled factors such as differences in body mass index), the results revealed consistent metabolic alterations in the plasma of lung cancer patients that are plausible within current knowledge of cancer metabolism. Indeed, our previous studies on lung tumor and noninvolved adjacent tissues pointed to changes in pathways common to those unveiled here, involving mainly glucose utilization, lipid metabolism, and the metabolism of free amino acids.^{32,33} It should also be noted that although alterations in these pathways have been reported for other cancer types, the particular set of metabolites found to be affected in this study (and their quantitative variations) may express a specific signature for lung cancer. Indeed, the possibility to simultaneously detect changes in multiple non-predefined metabolites constitutes an important advantage of metabonomics over conventional biochemical methods that assess individual compounds in a targeted manner. Together with the good discrimination achieved by NMR profiling of urine samples,³⁴ this work further supports the premise that lung cancer impacts systemic metabolism, thus opening new possibilities in terms of minimally invasive diagnostic methods. A particularly promising result in this respect was that the metabolic differences accounting for the discrimination between control and cancer groups were detectable right from stage I of the disease. Hence, the next step to determine the real clinical applicability of the NMR-based metabonomics approach will consist of testing the predictive ability of the constructed model on a new truly independent set of samples (collected from different subjects in another time frame). If this proves successful, a foreseen practical application for this work, which may contribute to improved lung cancer management, consists of using NMR metabolic profiling of blood plasma to identify subjects with higher probability of having asymptomatic lung cancer, within the general population or subgroups at increased risk (e.g., chronic smokers), in order to proceed with more advanced and specific radiological tests.

■ ASSOCIATED CONTENT

● Supporting Information

Supplementary Figures S1–S5. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +351 234401424. Fax: +351 234401470. E-mail: ioladuarte@ua.pt

■ ACKNOWLEDGMENT

Funding is acknowledged from the European Regional Development Fund through the Competitive Factors Thematic Operational Programme and from the Foundation for Science and Technology (FCT), Portugal (research project FCT/PTDC/QUI/68017/2006; FCOMP-01-0124-FEDER-007439). C.M.R. further acknowledges FCT for the doctoral grant SFRH/BD/63430/2009, and I.F.D. acknowledges L'Oréal Portugal, FCT, and the National UNESCO Committee for funding this work through the L'Oréal Medals of Honor for Women in Science 2007. The Portuguese National NMR Network (RNRMN), supported with FCT funds, is also acknowledged. The authors are grateful to all the volunteers who participated in this study, as well as to Dr. Jorge Ferreira and Eng. Nelson Ramos from INDASA S.A. for collaboration in the collection of control samples. Finally, Bruker BioSpin GmbH is acknowledged for access to BBIORFECODE.

■ REFERENCES

- (1) Boyle, P.; Levin, B. *World Cancer Report 2008*, 1st ed.; World Health Organization: Geneva, 2009; p 512.
- (2) Pirozynski, M. 100 years of lung cancer. *Resp. Med.* **2006**, *100* (12), 2073–2084.
- (3) Spratlin, J. L.; Serkova, N.; Eckhardt, S. G. Clinical applications of metabolomics in oncology: a review. *Clin. Cancer Res.* **2009**, *15* (2), 431–440.
- (4) Sitter, B.; Bathen, T. F.; Tessem, M.-B.; Gribbestad, I. S. High-resolution magic angle spinning (HR MAS) MR spectroscopy in metabolic characterization of human cancer. *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *54* (3–4), 239–254.
- (5) Keun, H. C.; Sidhu, J.; Pchejetski, D.; Lewis, J. S.; Marconell, H.; Patterson, M.; Bloom, S. R.; Amber, V.; Coombes, R. C.; Stebbing, J. Serum molecular signatures of weight change during early breast cancer chemotherapy. *Clin. Cancer Res.* **2009**, *15* (21), 6716–6723.
- (6) Asiago, V. M.; Alvarado, L. Z.; Shanaiah, N.; Gowda, G. A. N.; Owusu-Sarfo, K.; Ballas, R. A.; Raftery, D. Early detection of recurrent breast cancer using metabolite profiling. *Cancer Res.* **2010**, *70* (21), 8309–8318.
- (7) Gu, H. W.; Pan, Z. Z.; Xi, B. W.; Asiago, V.; Musselman, B.; Raftery, D. Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: Application to the detection of breast cancer. *Anal. Chim. Acta* **2011**, *686* (1–2), 57–63.
- (8) Gao, H.; Dong, B.; Liu, X.; Xuan, H.; Huang, Y.; Lin, D. Metabonomic profiling of renal cell carcinoma: High-resolution proton nuclear magnetic resonance spectroscopy of human serum with multivariate data analysis. *Anal. Chim. Acta* **2008**, *624* (2), 269–277.
- (9) Zira, A. N.; Theocharis, S. E.; Mitropoulos, D.; Migdalas, V.; Mikros, E. ¹H NMR metabonomic analysis in renal cell carcinoma: a possible diagnostic tool. *J. Proteome Res.* **2010**, *9* (8), 4038–4044.
- (10) Lin, L.; Yu, Q.; Yan, X.; Hang, W.; Zheng, J.; Xing, J.; Huang, B. Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabonomics? A serum metabonomic study of kidney cancer. *Analyst* **2010**, *135* (11), 2970–2978.
- (11) Xue, R.; Lin, Z.; Deng, C.; Dong, L.; Liu, T.; Wang, J.; Shen, X. A serum metabolomic investigation on hepatocellular carcinoma patients by chemical derivatization followed by gas chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom.* **2008**, *22* (19), 3061–3068.
- (12) Gao, H. C.; Lu, Q.; Liu, X.; Cong, H.; Zhao, L. C.; Wang, H. M.; Lin, D. H. Application of H-1 NMR-based metabonomics in the study of metabolic profiling of human hepatocellular carcinoma and liver cirrhosis. *Cancer Sci.* **2009**, *100* (4), 782–785.
- (13) Yin, P.; Wan, D.; Zhao, C.; Chen, J.; Zhao, X.; Wang, W.; Lu, X.; Yang, S.; Gu, J.; Xu, G. A metabonomic study of hepatitis B-induced liver cirrhosis and hepatocellular carcinoma by using RP-LC and HILIC coupled with mass spectrometry. *Mol. BioSyst.* **2009**, *5* (8), 868–876.
- (14) Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X.; Byun, J.; S. Omenn, G.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M. Metabonomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009**, *457* (7231), 910–914.
- (15) Lokhov, P. G.; Dashtiev, M. I.; Bondartsov, L. V.; Lisitsa, A. V.; Moshkovskii, S. A.; Archakov, A. I. Metabolic fingerprinting of blood plasma from patients with prostate cancer. *Biochemistry (Moscow) Supp. Series B: Biomed. Chem.* **2010**, *4* (1), 37–41.
- (16) Thysell, E.; Surowiec, I.; Hornberg, E.; Crnalic, S.; Widmark, A.; Johansson, A. L.; Stattin, P.; Bergh, A.; Moritz, T.; Antti, H.; Wikstro, P. Metabolomic characterization of human prostate cancer bone metastases reveals increased levels of cholesterol. *PLoS ONE* **2010**, *5* (11), e14175.
- (17) Ludwig, C.; Ward, D. G.; Martin, A.; Viant, M. R.; Ismail, T.; Johnson, P. J.; Wakelam, M. J. O.; Gunther, U. L. Fast targeted multi-dimensional NMR metabolomics of colorectal cancer. *Magn. Reson. Chem.* **2009**, *47* (S1), S68–S73.
- (18) Qiu, Y. P.; Cai, G. X.; Su, M. M.; Chen, T. L.; Zheng, X. J.; Xu, Y.; Ni, Y.; Zhao, A. H.; Xu, L. X.; Cai, S. J.; Jia, W. Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS. *J. Proteome Res.* **2009**, *8* (10), 4844–4850.
- (19) Ritchie, S. A.; Ahiahonu, P. W.; Jayasinghe, D.; Heath, D.; Liu, J.; Lu, Y.; Jin, W.; Kavianpour, A.; Yamazaki, Y.; Khan, A. M.; Hossain, M.; Su-Myat, K. K.; Wood, P. L.; Krenitsky, K.; Takemasa, L.; Miyake, M.; Sekimoto, M.; Monden, M.; Matsubara, H.; Nomura, F.; Goodenowe, D. B. Reduced levels of hydroxylated, polyunsaturated ultra long-chain fatty acids in the serum of colorectal cancer patients: implications for early screening and detection. *BMC Med.* **2010**, *8* (13).
- (20) Ma, Y.; Liu, W.; Peng, J.; Huang, L.; Zhang, P.; Zhao, X.; Cheng, Y.; Qin, H. A pilot study of gas chromatography/mass spectrometry-based serum metabolic profiling of colorectal cancer after operation. *Mol. Biol. Rep.* **2010**, *37* (3), 1403–1411.
- (21) Tiziani, S.; Lopes, V.; Günther, U. L. Early stage diagnosis of oral cancer using ¹H NMR-based metabolomics. *Neoplasia* **2009**, *11* (3), 269–276.
- (22) Zhou, J.; Xu, B.; Huang, J.; Jia, X.; Xue, J.; Shi, X.; Xiao, L.; Li, W. ¹H NMR-based metabonomic and pattern recognition analysis for detection of oral squamous cell carcinoma. *Clin. Chim. Acta* **2009**, *401* (1–2), 8–13.
- (23) Urayama, S.; Zou, W.; Brooks, K.; Tolstikov, V. Comprehensive mass spectrometry based metabolic profiling of blood plasma reveals potent discriminatory classifiers of pancreatic cancer. *Rapid Commun. Mass Spectrom.* **2010**, *24* (5), 613–620.
- (24) Bathe, O. F.; Shaykhtudinov, R.; Kopciuk, K.; Weljie, A. M.; McKay, A.; Sutherland, F. R.; Dixon, E.; Dunse, N.; Sotiropoulos, D.; Vogel, H. J. Feasibility of identifying pancreatic cancer based on serum metabolomics. *Cancer Epidemiol. Biomarkers Prev.* **2011**, *20* (1), 140–147.
- (25) Zhang, J. A.; Liu, L. Y.; Wei, S. W.; Gowda, G. A. N.; Hammoud, Z.; Kesler, K. A.; Raftery, D. Metabolomics study of esophageal adenocarcinoma. *J. Thorac. Cardiovasc. Surg.* **2011**, *141* (2), 469–U1048.

- (26) Zhang, Z. Y.; Qiu, Y. P.; Hua, Y. Q.; Wang, Y. H.; Chen, T. L.; Zhao, A. H.; Chi, Y.; Pan, L.; Hu, S.; Li, J. A.; Yang, C. W.; Li, G. D.; Sun, W.; Cai, Z. D.; Jia, W. Serum and urinary metabolomic study of human osteosarcoma. *J. Proteome Res.* **2010**, *9* (9), 4861–4868.
- (27) Maeda, J.; Higashiyama, M.; Imaizumi, A.; Nakayama, T.; Yamamoto, H.; Daimon, T.; Yamakado, M.; Imamura, F.; Kodama, K. 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* **2010**, *10* (690).
- (28) Dong, J.; Cai, X. M.; Zhao, L. L.; Xue, X. Y.; Zou, L. J.; Zhang, X. L.; Liang, X. M. Lysophosphatidylcholine profiling of plasma: discrimination of isomers and discovery of lung cancer biomarkers. *Metabolomics* **2010**, *6* (4), 478–488.
- (29) Jordan, K. W.; Adkins, C. B.; Su, L.; Halpern, E. F.; Mark, E. J.; Christiani, D. C.; Cheng, L. L. Comparison of squamous cell carcinoma and adenocarcinoma of the lung by metabolomic analysis of tissue–serum pairs. *Lung Cancer* **2009**, *68* (1), 44–50.
- (30) Laiakis, E. C.; Morris, G. A. J.; Fornace, A. J.; Howie, S. R. C. Metabolomic analysis in severe childhood pneumonia in the gambia, West Africa: findings from a pilot study. *PLoS ONE* **2010**, *5* (9), e12655.
- (31) Stringer, K. A.; Serkova, N. J.; Karnovsky, A.; Guire, K.; Paine, R.; Standiford, T. J. Metabolic consequences of sepsis-induced acute lung injury revealed by plasma H-1-nuclear magnetic resonance quantitative metabolomics and computational analysis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2011**, *300* (1), L4–L11.
- (32) Rocha, C. M.; Barros, A. S.; Gil, A. M.; Goodfellow, B. J.; Humpfer, E.; Spraul, M.; Carreira, I. M.; Melo, J. B.; Bernardo, J.; Gomes, A.; Sousa, V.; Carvalho, L.; Duarte, I. F. Metabolic profiling of human lung cancer tissue by ¹H high resolution magic angle spinning (HRMAS) NMR spectroscopy. *J. Proteome Res.* **2010**, *9* (1), 319–332.
- (33) Duarte, I. F.; Rocha, C. M.; Barros, A. S.; Gil, A. M.; Goodfellow, B. J.; Carreira, I. M.; Bernardo, J.; Gomes, A.; Sousa, V.; Carvalho, L. Can nuclear magnetic resonance (NMR) spectroscopy reveal different metabolic signatures for lung tumors? *Virchows Archiv.* **2010**, *457* (6), 715–725.
- (34) Carrola, J.; Rocha, C. M.; Barros, A. S.; Gil, A. M.; Goodfellow, B. J.; Carreira, I. M.; Bernardo, J.; Gomes, A.; Sousa, V.; Carvalho, L.; Duarte, I. F. Metabolic signatures of lung cancer in biofluids: NMR-based metabolomics of urine. *J. Proteome Res.* **2011**, *10* (1), 221–230.
- (35) Travis, W. D.; Brambilla, E.; Müller-Hermelink, H. K.; Harris, C. *World Health Organization Classification of Tumours; Tumours of Lung, Pleura, Thymus and Heart*; IARC Press: Lyon, France, 2004.
- (36) Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S. The human serum metabolome. *PLoS ONE* **2011**, *6* (2), e16957.
- (37) Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002**, *16* (3), 119–128.
- (38) Xu, Q. S.; Liang, Y. Z. Monte Carlo cross validation. *Chemom. Intell. Lab. Syst.* **2001**, *56* (1), 1–11.
- (39) Duarte, I. F.; Goodfellow, B. J.; Barros, A.; Jones, J. G.; Barosa, C.; Diogo, L.; Garcia, P.; Gil, A. M. Metabolic characterization of plasma in juveniles with glycogen storage disease type 1a (GSD1a) by high resolution ¹H NMR spectroscopy. *NMR Biomed.* **2007**, *20* (4), 401–412.
- (40) Daykin, C. A.; Corcoran, O.; Hansen, S. H.; Bjørnsdottir, I.; Cornett, C.; Connor, S. C.; Lindon, J. C.; Nicholson, J. K. Application of directly coupled HPLC NMR to separation and characterization of lipoproteins from human serum. *Anal. Chem.* **2001**, *73* (6), 1084–1090.
- (41) Liu, M.; Tang, H.; Nicholson, J. K.; Lindon, J. C. Use of ¹H NMR-determined diffusion coefficients to characterize lipoprotein fractions in human blood plasma. *Magn. Reson. Chem.* **2002**, *40* (13), S83–S88.
- (42) Bell, J. D.; Brown, J. C. C.; Nicholson, J. K.; Sadler, P. J. Assignment of resonances for ‘acute-phase’ glycoproteins in high resolution proton NMR spectra of human blood plasma. *FEBS Lett.* **1987**, *215* (2), 311–315.
- (43) Fiorenza, A. M.; Branchi, A.; Sommariva, D. Serum lipoprotein profile in patients with cancer. A comparison with non-cancer subjects. *Int. J. Clin. Lab. Res.* **2000**, *30* (3), 141–145.
- (44) Muntoni, S.; Atzori, L.; Mereu, R.; Satta, G.; Macis, M. D.; Congia, M.; Tedde, A.; Desogus, A.; Muntoni, S. Serum lipoproteins and cancer. *Nutr. Metab. Cardiovas.* **2009**, *19* (3), 218–225.
- (45) Fiorenza, A. M.; Branchi, A.; Cardena, A.; Molgora, M.; Rovellini, A.; Sommariva, D. Serum cholesterol levels in patients with cancer. Relationship with nutritional status. *Int. J. Clin. Lab. Res.* **1996**, *26* (1), 37–42.
- (46) Fan, T. W. M.; Lane, A. N.; Higashi, R. M.; Farag, M. A.; Gao, H.; Bousamra, M.; Miller, D. M. Altered regulation of metabolic pathways in human lung cancer discerned by ¹³C stable isotope-resolved metabolomics (SIRM). *Mol. Cancer* **2009**, *8* (41).
- (47) MacIntyre, D. A.; Jiménez, B.; Lewintre, E. J.; Martín, C. R.; Schäfer, H.; MBallesteros, C. G.; Mayans, J. R.; Spraul, M.; García-Conde, J.; Pineda-Lucena, A. Serum metabolome analysis by ¹H-NMR reveals differences between chronic lymphocytic leukaemia molecular subgroups. *Leukemia* **2010**, *24* (4), 788–797.
- (48) Erol, A. Retrograde regulation due to mitochondrial dysfunction may be an important mechanism for carcinogenesis. *Med. Hypotheses* **2005**, *65* (3), 525–529.
- (49) Brière, J.-J.; Favier, J.; Gimenez-Roqueplo, A.-P.; Rustin, P. Tricarboxylic acid cycle dysfunction as a cause of human diseases and tumor formation. *Am. J. Physiol. Cell Physiol.* **2006**, *291* (6), C1114–C1120.
- (50) Lai, H.-S.; Lee, J.-C.; Lee, P.-H.; Wang, S.-T.; Chen, W.-J. Plasma free amino acid profile in cancer patients. *Semin. Cancer Biol.* **2005**, *15* (4), 267–276.
- (51) Heber, D.; Byerly, L. O.; Chlebowski, R. T. Metabolic abnormalities in the cancer patient. *Cancer* **1985**, *11* (1), 225–259.
- (52) Naini, A. B.; Dickerson, J. W.; Brown, M. M. Preoperative and postoperative levels of plasma protein and amino acid in esophageal and lung cancer patients. *Cancer* **1988**, *62* (2), 355–360.
- (53) Norton, J. A.; Gorschboth, C. M.; Wesley, R. A.; Burt, M. E.; Brennan, M. F. Fasting plasma amino acid levels in cancer patients. *Cancer* **1985**, *56* (5), 1181–1186.
- (54) Okamoto, N.; Miyagi, Y.; Chiba, A.; Akaike, M.; Shiozawa, M.; Imaizumi, A.; Yamamoto, H.; Ando, T.; Yamakado, M.; Tochikubo, O. Diagnostic modeling with differences in plasma amino acid profiles between non-cachectic colorectal/breast cancer patients and healthy individuals. *Int. J. Med. Med. Sci.* **2009**, *1* (1), 1–8.
- (55) Haffner, H. T.; Graw, M.; Besserer, K.; Blicke, U.; Henge, C. Endogenous methanol: variability in concentration and rate of production. Evidence of a deep compartment? *Forensic Sci. Int.* **1996**, *79* (2), 145–154.
- (56) Freedman, D. S.; Otvos, J. D.; Jeyarajah, E. J.; Shalurova, I.; Cupples, L. A.; Parise, H.; D’Agostino, R. B.; Wilson, P. W.; Schaefer, E. J. Sex and age differences in lipoprotein subclasses measured by nuclear magnetic resonance spectroscopy: the Framingham Study. *Clin. Chem.* **2004**, *50* (7), 1189–1200.
- (57) Kochhar, S.; Jacobs, D. M.; Ramadan, Z.; Berruex, F.; Fuerholz, A.; Fay, L. B. Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabolomics. *Anal. Biochem.* **2006**, *352* (2), 274–281.